M N 48

Eine Iterative Methode zur Lösung linearer Gleichungssysteme und Fehlerabschätzungen zum Einzelschrittverfahren

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Hamburg

vorgelegt von
Harry Feldmann
aus Hamburg

Hamburg
1962

Eine Iterative Methode zur Lösung linearer Gleichungssysteme und Fehlerabschätzungen zum Einzelschrittverfahren

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Hamburg

vorgelegt von
Harry Feldmann
aus Hamburg

Hamburg
1962

Genehmigt von der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Hamburg
auf Antrag von Professor Dr. Lothar Collatz

Hamburg, den 10. Februar 1962

Professor Dr. Curt Kosswig
Dekan

R. UNIV. BIBLIOTHEEK LEIDEN

Inhaltsverzeichnis

Seite

§ 1. Einleitung

1. Direkte und Iterative Methoden zur Lösung größerer linearer Gleichungssysteme 5

2. Übersicht und Problemstellung 8

§ 2. Theorie der Overrelaxation nach D. YOUNG 11

§ 3. Iterative Bestimmung von Eigenwerten und Konvergenzbeschleunigung nach THIELE, AITKEN, WYNN sowie neue Formeln 17

§ 4. Ein Satz über die Darstellung der Fehlervektoren bei der Overrelaxation 29

§ 5. Iterative Bestimmung des optimalen Relaxationsfaktors ω_b und Konvergenzbeschleunigung bei der Overrelaxation 33

§ 6. Fehlerabschätzungen zum Einzelschrittverfahren

1. Hilfsmittel 39

2. Ein neues hinreichendes Konvergenzkriterium und eine neue Fehlerabschätzung 41

3. Verallgemeinerungen der Fehlerabschätzungen von L. COLLATZ und W. DÜCK 42

4. Ein neues hinreichendes Konvergenzkriterium und eine neue Fehlerabschätzung bei Matrizen A "von strenger property (A)" 44

5. Vergleich der Konvergenzkriterien und Fehlerabschätzungen. 46

§ 7. Beispiele 48

LITERATURVERZEICHNIS 62

§ 1. EINLEITUNG

Die praktische Behandlung vieler Probleme aus Anwendungsbereichen, wie zum Beispiel

- Differenzenverfahren bei gewöhnlichen und partiellen Differentialgleichungen und Integralgleichungen,
- Approximation und Ausgleichsrechnung nach der Methode der kleinsten mittleren Fehlerquadrate,
- Berechnung elektrischer Netzwerke,
- Gleichgewichtsbestimmung in der Statik,
- Vermessungskunde,

und anderer mehr führt auf die Lösung großer Systeme linearer Gleichungen. Mit dem Einsatz moderner elektronischer Rechenanlagen ist es möglich geworden, Systeme von Hunderten bis Tausenden von Gleichungen zu lösen. Dazu benötigt man Methoden, die wenig Rechenaufwand und daher ein erträgliches Maß an Rechenzeit erfordern, die "stabil" [10] gegen Verfälschung der Lösung durch auftretende Rundungsfehler und der begrenzten Speicherkapazität der Maschine angepaßt sind.

1. Direkte und Iterative Methoden zur Lösung größerer linearer Gleichungssysteme.

Ein großer Teil der mathematischen Literatur ist den Methoden zur Lösung linearer Gleichungssysteme gewidmet, und doch sind bis heute viele wichtige und interessante Fragen offen geblieben. Man unterteilt in 'Direkte Methoden' und 'Iterative Methoden'.

Direkte Methoden werden so genannt, weil sie die exakte Lösung in einer vorher bestimmten endlichen Anzahl von Multiplikationen

und Additionen liefern. Diese Anzahl und damit der Rechenaufwand wächst bei den gebräuchlichen Verfahren mit der dritten Potenz n^3 der Anzahl n der Gleichungen; beim GAUSS'schen Eliminationsverfahren zum Beispiel benötigt man $\frac{1}{3}n(n^2+3n-1)$ Multiplikationen und $\frac{1}{6}n(n-1)(2n+5)$ Additionen. Im gleichen Maße (n^3) wächst bei den Direkten Methoden der Einfluß der Rundungsfehler, die benötigte Speicherkapazität wächst etwa mit n^2.

Gegenwärtig wird es als unpraktisch angesehen [16], größere als Systeme von 1000 Gleichungen direkt zu lösen.

Iterative Methoden beruhen auf der wiederholten Anwendung eines Algorithmus, der eine Näherung durch eine andere, in gewissem Sinne bessere Näherung für die wahre Lösung, ersetzt. Konvergiert das Verfahren, so erhält man - abgesehen von Rundungsfehlern - nach hinreichend vielen Schritten eine Lösung, die der exakten Lösung beliebig nahe kommt. Ein Iterationsschritt besteht bei diesen Methoden im wesentlichen aus einer Matrix × Vektor-Multiplikation und erfordert einen Rechenaufwand proportional n^2. Der Rechenaufwand der Iterativen Methode ist daher proportional $m.n^2$, wenn m die Anzahl der Iterationen ist, die zur Berechnung der wahren Lösung (auf gewünschte Stellenzahl) nötig ist. Zum Beispiel benötigte D. YOUNG [26] mit Overrelaxation zur Lösung spezieller Gleichungssysteme (auf 5 gültige Stellen) bei n = 400 Gleichungen m = 35 und bei n = 90 000 Gleichungen m = 640 Iterationsschritte. Der Einfluß der Rundungsfehler wächst, da die vorgegebene Koeffizientenmatrix bei jeder Iteration wieder unverfälscht benutzt wird, nur

mit n^2 und kann durch Mitführung von Schutzstellen bei der Rechnung ausgeglichen werden. Die benötigte Speicherkapazität ist proportional n^2, sie kann in Spezialfällen, falls der gegebenen Koeffizientenmatrix ein einfach zu programmierendes Bildungsgesetz zugrunde liegt (Differenzenverfahren in speziellen Bereichen), und auf Grund des einfach zu programmierenden Iterations-Algorithmus, auf etwa 2n bis 3n Speicher herabgedrückt werden.

Die bekanntesten Iterativen Methoden zur Lösung linearer Gleichungssysteme werden in der folgenden Zusammenstellung genannt:

Ein System

$$Bx - r = 0 \quad \ldots\ldots\ldots\ldots \quad (1,1) \,,$$

dessen Koeffizientenmatrix ('durchdividierte' Form)

$B = -A_L + E - A_R$ sich zusammensetzt aus einer linken unteren Dreiecksmatrix A_L (Elemente a_{jk} für $j > k$, sonst 0), einer rechten oberen Dreiecksmatrix A_R (Elemente a_{jk} für $j < k$, sonst 0) und der Einheitsmatrix E, wird in einem Schritt exakt gelöst durch das ('Newtonsche') Iterationsverfahren (m=0, 1, 2, ...)

$$x_{m+1} = x_m - B^{-1}(Bx_m - r) \quad \ldots\ldots\ldots\ldots \quad (1,2) \,.$$

Ersetzt man in (1, 2) die i.a. nicht bekannte Matrix

$$B^{-1} = (-A_L + E - A_R)^{-1} \quad \ldots\ldots\ldots \quad (1,3)$$

näherungsweise durch die Matrix

$$(-\vartheta A_L + \frac{1}{\omega_m} E)^{-1} \quad \ldots\ldots\ldots\ldots \quad (1,4) \,,$$

so erhält man folgende ('Vereinfachte Newtonsche') Iterationsverfahren:

δ	ω_m		
1	1	Einzelschrittverfahren	C.F.GAUSS [12] L.SEIDEL [21]
1	ω $(0<\omega<2)$	Overrelaxation	R.V.SOUTHWELL [22], D.YOUNG [26]
0	1	Gesamtschrittverfahren	J.LIOUVILLE 1838 C.G.J.JAKOBI [15]
0	$\omega_m^{-1} = \frac{d_m d_m}{d_m B d_m}$	Gradientenmethode (B symm. pos. def.) $d_m = Bx_m - r$	G.TEMPLE [24]
0	ω_m^{-1}=(m+1)te Nullstelle von $P_k(x)=\frac{T_k(\frac{b+a-2x}{b-a})}{T_k(\frac{b+a}{b-a})}$	RICHARDSONS Methode (B symm. pos. def. Eigenwerte in $<a,b>$) $T_k(x)=\cos(k \arccos x)$ $m<k$, $0<a<b$	L.F. RICHARDSON [18]

In neuerer Zeit sind aus diesen Prototypen Iterativer Methoden weitere Verfahren entwickelt worden, z.B. aus der Overrelaxation die 'Block-Iteration' von ARMS, GATES, ZONDEK [4], für das Gesamtschrittverfahren Konvergenzbeschleunigung von GOLUB, VARGA [13], aus der Gradientenmethode die 'Methode der konjugierten Gradienten' von HESTENES, STIEFEL [14] und aus RICHARDSONS Methode Verfahren höherer Ordnung von ARMS, GATES, ZONDEK [4], E.L.STIEFEL [23] u.a.

2. Übersicht und Problemstellung.

Diese Arbeit bringt Ergänzungen zur Theorie der Overrelaxation, die sich in der Praxis - insbesondere bei größeren Gleichungssystemen - im Vergleich zu anderen Iterativen Methoden bisher gut bewährt hat [16]. Für den Spezialfall $\omega = 1$ (Einzelschrittverfahren) werden neue Fehlerabschätzungen gewonnen.

In § 2 findet man die Theorie der Overrelaxation von D.YOUNG [26] für Matrizen $A = A_L + A_R$ 'von strenger property (A)'. Die

Bestimmung des optimalen Relaxationsfaktors ω_b aus dem betragsmäßig größten Eigenwert der Matrix $L(\omega)$ und die Berechnung der Eigenwerte von $L(\omega)$ aus den Eigenwerten von A wird unter Verwendung neuerer Darstellungen von B. FRIEDMAN [11] und FORSYTHE, WASOW [10] hergeleitet.

§ 3 behandelt allgemeine Iterationsverfahren $x_{m+1} = Lx_m + s$ $(m = 0, 1, 2, \ldots)$, deren Fehlervektor ϵ_m sich mit den Eigenwerten t_ν $(\nu = 1, \ldots, n)$ von L und Vektoren $c_\nu \neq 0$ darstellen läßt als $\epsilon_m = \sum_{\nu=1}^{n} c_\nu t_\nu^m$. Bei diesen Verfahren ist, falls $|t_1| > |t_2| \geq |t_3| \geq \ldots \geq |t_n|$, die iterative Bestimmung zumindest des ersten Eigenwertes t_1 nach dem δ^2-Verfahren von A. C. AITKEN [1] und Konvergenzbeschleunigung nach THIELE (Interpolationsrechnung 1909), AITKEN [2] möglich. Es werden neue Formeln zur Bestimmung höherer Eigenwerte und neue Formeln für Konvergenzbeschleunigung mit Restgliedern höherer Ordnung hergeleitet.

In § 4 wird der neue Satz bewiesen, daß die Overrelaxation mit symmetrischen Matrizen A 'von strenger property (A)' ein Iterationsverfahren der in § 3 definierten Art ist.

In § 5 können daher die Methoden von § 3 auf die Overrelaxation angewandt werden. Es wird eine iterative Methode ausgearbeitet, die ohne vorherige Kenntnis der Eigenwerte von $L(\omega)$ oder A und daher ohne vorherige Kenntnis des optimalen Faktors ω_b so schnell konvergiert wie Overrelaxation mit als bekannt vorausgesetztem besten Faktor ω_b.

In § 6 findet man zwei neue Fehlerabschätzungen zum Einzelschrittverfahren ($\omega = 1$), beide formuliert für beliebige Matrixnormen, von denen die erste, H. FELDMANN [9], gültig für beliebige Matrizen $A = A_L + A_R$, die Dreiecksform der Matrizen A_L, A_R und die zweite, gültig für Matrizen $A = A_L + A_R$ 'von strenger property (A)', die spezielle Rechtecksform der Matrizen A_L, A_R benutzt.

Außerdem werden zwei neue hinreichende Konvergenzkriterien sowie allgemeiner als bisher die Fehlerabschätzungen von W. DÜCK [7] für beliebige und die Fehlerabschätzung von L. COLLATZ [5] für monotone absolute Matrixnormen hergeleitet.

In § 7 sind Beispiele zu den §§ 1-6 zusammengestellt. Es wird gezeigt, daß Differenzenverfahren bei 1. Randwertaufgaben elliptischer Differentialgleichungen auf lineare Gleichungssysteme mit symmetrischen Matrizen A 'von strenger property (A)' führen. Mit Ausnahme der numerisch einfachen Fehlerabschätzungen von § 6 wurden die Beispiele auf der elektronischen Rechenanlage IBM 650 (2000 Speicher) des Instituts für Angewandte Mathematik der Universität Hamburg programmiert und gerechnet.

§ 2. THEORIE DER OVERRELAXATION NACH D. YOUNG .

Wir führen eine Vektornorm $\| x \|$ [8] ein und betrachten Iterationen im BANACHschen Raum R_n der Vektoren x.

Zur Lösung eines linearen Gleichungssystems

$$x = Ax + r \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2,1) \ ,$$

dessen $n \times n$ Koeffizientenmatrix $A = A_L + A_R$ als Summe einer linken unteren Dreiecksmatrix A_L (mit Elementen a_{jk}, $j > k$, und sonst 0) und einer rechten oberen Dreiecksmatrix A_R (mit Elementen a_{jk}, $j < k$, und sonst 0) gegeben ist, wählt man einen beliebigen Anfangsvektor x_0 und bringt in zeilenweiser 'Over'-Relaxation bei der i-ten Komponente das ω-fache der (additiven) Änderung an, die nötig wäre, um die i-te Gleichung exakt zu erfüllen (i=1, 2, ..., n).

Relaxiert man von x_m nach x_{m+1}, so ist diese Änderung bis auf entgegengesetztes Vorzeichen gleich der i-ten Komponente des 'Defektes'

$$-A_L x_{m+1} + E x_m - A_R x_m - r \quad ,$$

und die Iterationsvorschrift für die (komponentenweise, i=1, 2, ..., n) Overrelaxation lautet demnach

$$x_{m+1} = x_m + \omega (A_L x_{m+1} - E x_m + A_R x_m + r) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2,2)$$

$$m = (0, 1, 2, \ldots .) .$$

Daraus erhält man durch Umformung die Darstellung

$$x_{m+1} = (E - \omega A_L)^{-1} ((1-\omega) E + \omega A_R) x_m + (E - \omega A_L)^{-1} \omega r \quad . \quad . \quad . \quad . \quad . \quad (2,3) .$$

Nach einem Satz von R. OLDENBURGER [17] konvergiert die Overrelaxation (2, 2) genau dann gegen die dann eindeutig existierende Lösung von (2, 1), wenn der Spekatralradius $\varrho (L(\omega))$,

d.h. das Maximum der Beträge der charakteristischen Zahlen (Eigenwerte) von

$$L(\omega) = (E-\omega A_L)^{-1}((1-\omega)E+\omega A_R) \quad \ldots\ldots\ldots\ldots \quad (2,4)$$

kleiner als 1 ist.

Da für große m im Sinne irgendeiner Vektornorm der Quotient $\|\epsilon_{m+1}\| / \|\epsilon_m\|$ der Fehler $\epsilon_m = x - x_m$ gegen den Spektralradius $\varrho(L(\omega))$ strebt (siehe etwa FORSYTHE, WASOW [10]), ist der bestmögliche Wert ω_b für den Relaxationsfaktor durch die Forderung

$$\varrho(L(\omega)) = \text{Min} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2,5)$$

bestimmt.

Für spezielle Matrizen A von 'property (A)' hat D. YOUNG [26] die Eigenwerte $\eta_i(\omega)$ $(i=1,\ldots,n)$ der Matrix $L(\omega)$ und den Faktor ω_b in Abhängigkeit von den Eigenwerten λ_i der Matrix A bestimmen können.

Definition : Eine $n \times n$ Matrix A heißt 'von strenger property (A)', wenn sie zwei Nullmatrizen ($n_1 \times n_1$ und $n_2 \times n_2$, $n_1 + n_2 = n$) als Diagonalblöcke besitzt:

$$\begin{array}{c} \\ n_1 \\ n_2 \end{array} \begin{array}{c} \overset{n_1}{\longleftrightarrow}\ \overset{n_2}{\longleftrightarrow} \\ \left(\begin{array}{c|c} 0 & A_R \\ \hline A_L & 0 \end{array}\right) \end{array} \quad \ldots\ldots\ldots\ldots \quad (2,6)$$

Setzt man die Matrix A des Gleichungssystems (2,1) als 'von strenger property (A)' voraus, so kann die (komponentenweise) Overrelaxation (2,2) nicht nur in der Reihenfolge $i=1,2,\ldots,n$, sondern in beliebiger Reihenfolge der ersten n_1 Komponenten und anschließend in beliebiger Reihenfolge der übrigen n_2 Kom-

durchgeführt werden. Vertauschung der beiden Indizesmengen hat Vertauschung von A_L und A_R in Formel (2,2) zur Folge ('Overrelaxation von unten nach oben').

Die im Hinblick auf § 4 gegebene Definition (2,6) 'von strenger property (A)' ist spezieller als die YOUNGsche [26] Definition 'von property (A)', aber eine Matrix A 'von property (A)' des Systems (2,1) kann stets durch identische Permutationen von Zeilen und Spalten auf eine Gestalt 'von strenger property (A)' gebracht werden [10].

Die genannten möglichen Reihenfolgen der (komponentenweisen) Overrelaxation entsprechen den nach D. YOUNG [26] zulässigen 'consistent orderings'.

Ist A 'von strenger property (A)', so folgt für das zugehörige charakteristische Polynom

$$
\begin{aligned}
P(\lambda) &= \det(A_L + A_R - \lambda E) \\
&= \det\begin{pmatrix} -\lambda E_1 & A_R \\ A_L & -\lambda E_2 \end{pmatrix} \\
&= \det\left[\begin{pmatrix} -E_1 & 0 \\ 0 & E_2 \end{pmatrix}\cdot\begin{pmatrix} -\lambda E_1 & A_R \\ A_L & -\lambda E_2 \end{pmatrix}\cdot\begin{pmatrix} -E_1 & 0 \\ 0 & E_2 \end{pmatrix}\right] \qquad \cdots\cdots\cdots \quad (2,7) \\
&= \det\begin{pmatrix} -\lambda E_1 & -A_R \\ -A_L & -\lambda E_2 \end{pmatrix} \\
&= (-1)^n P(-\lambda) \quad ,
\end{aligned}
$$

d.h. A besitzt p Paare von Eigenwerten $\pm\lambda_j \neq 0$ $(j=1,\ldots,p)$ und $n-2p$ Eigenwerte 0 $(2p \leq n)$.

Durch Vergleich mit dem zu $L(\omega)$ gehörigen charakteristischen Polynom $(\omega \neq 0)$:

$$\det\left[(E-\omega A_L)^{-1}((1-\omega)E+\omega A_R) - \eta E\right] = 0$$

bzw. $$Q(\eta) = \det\left[\eta A_L + \omega^{-1}(1-\omega-\eta)E + A_R\right] = 0$$

$$= \det \begin{pmatrix} \omega^{-1}(1-\omega-\eta)E_1 & A_R \\ \eta A_L & \omega^{-1}(1-\omega-\eta)E_2 \end{pmatrix} \quad \ldots\ldots \quad (2,8)$$

$$= \eta^{\frac{n}{2}} \det \left[\begin{pmatrix} E_1 & 0 \\ 0 & \eta^{-\frac{1}{2}} E_2 \end{pmatrix} \cdot \begin{pmatrix} \omega^{-1}(1-\omega-\eta)E_1 & A_R \\ \eta A_L & \omega^{-1}(1-\omega-\eta)E_2 \end{pmatrix} \cdot \begin{pmatrix} \eta^{-\frac{1}{2}} E_1 & 0 \\ 0 & E_2 \end{pmatrix} \right]$$

$$= \eta^{\frac{n}{2}} \det \begin{pmatrix} \eta^{-\frac{1}{2}} \omega^{-1}(1-\omega-\eta)E_1 & A_R \\ A_L & \eta^{-\frac{1}{2}} \omega^{-1}(1-\omega-\eta)E_2 \end{pmatrix}$$

$$= \eta^{\frac{n}{2}} \, P(\eta^{-\frac{1}{2}} \omega^{-1}(-1+\omega+\eta))$$

sieht man, daß zu jedem der p Paare von Eigenwerten $\pm \lambda_j \neq 0$ von A ein Paar von Eigenwerten η_j von $L(\omega)$ mit

$$\eta_j^{-\frac{1}{2}} \omega^{-1}(-1+\omega+\eta_j) = \pm \lambda_j \quad \ldots\ldots\ldots \quad (2,9)$$

bzw. $$\eta_j^{\frac{1}{2}} = \frac{\omega\lambda_j}{2} \pm \frac{1}{2} \sqrt{\omega^2 \lambda_j^2 - 4(\omega-1)} \quad \ldots\ldots\ldots \quad (2,10)$$

bzw. $$\eta_j = \frac{\omega^2 \lambda_j^2 - 2(\omega-1)}{2} \pm \frac{\omega\lambda_j}{2} \sqrt{\omega^2 \lambda_j^2 - 4(\omega-1)} \quad \ldots \quad (2,11)$$

und zu jedem der n-2p Eigenwerte 0 von A ein Eigenwert $1-\omega$ von $L(\omega)$ zugehört $(\omega \neq 0)$.

Speziell für $\omega = 1$ (Einzelschrittverfahren) gehört zu jedem der p Paare von Eigenwerte $\pm \lambda_j \neq 0$ von A ein Paar von Eigenwerten $\eta_j = 0, \lambda_j^2$ von $L(\omega)$ und zu jedem der n-2p Eigenwerte 0 von A ein Eigenwert 0 von $L(\omega)$.

Die Transformation (2,9) setzt sich im allgemeinen Fall komplexer $\lambda_j, \eta_j, \omega$ $(\omega \neq 0)$ zusammen aus den konformen Abbildungen

$$\lambda_j = \frac{1}{\omega} \left(z_j + (\omega-1) \frac{1}{z_j} \right) \qquad \text{(JOUKOWSKI)}$$

und $$z_j = \sqrt{\eta_j} \quad .$$

Satz (2,11): Beschränkt man sich auf **reelle Relaxationsfaktoren** ω und auf Matrizen A 'von strenger property (A)' mit **reellen Eigenwerten** λ_i $(i=1,\dots,n)$, so konvergiert die Overrelaxation (2,2) genau für

$$-1 < \lambda_i < +1 \qquad (i=1,\dots,n)$$

und $$0 < \omega < 2$$

gegen die dann eindeutig existierende Lösung von (2,1).

Beweis: Wie die Diskussion von (2,11) in Abb. 1 zeigt, sind für $-1 < \lambda_i < +1$ genau dann alle Eigenwerte $\eta_i(\omega)$ vom Betrage kleiner als 1, wenn $0 < \omega < 2$ ist. Im Falle $|\lambda_j| > 1$ aber gilt nach (2,10) für einen der beiden Werte $\eta_j^{\frac{1}{2}}$ und beliebig reelles ω

$$|\eta_j^{\frac{1}{2}}| = \tfrac{1}{2}|\omega\lambda_j| + \tfrac{1}{2}\sqrt{(\omega-2)^2 + \omega^2(\lambda_j^2-1)}$$

$$> \tfrac{1}{2}|\omega| + \tfrac{1}{2}|\omega-2| \geq 1 \qquad \text{(q.e.d.)}$$

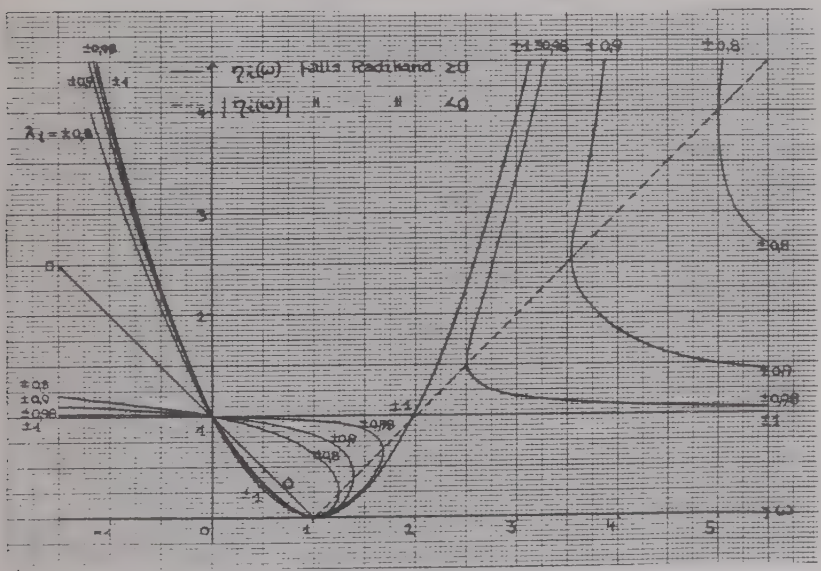


Abb. 1

$$\eta_i(\omega) = \frac{\omega^2\lambda_i^2 - 2(\omega-1)}{2} \pm \frac{\omega\lambda_i}{2}\sqrt{\omega^2\lambda_i^2 - 4(\omega-1)} \quad \dots\dots (2,11)$$

in Abhängigkeit vom Parameter λ_i

Den optimalen Relaxationsfaktor ω_b bestimmt man aus der Forderung (2,5), d.h. man sucht im Intervall $0 < \omega < 2$ das Betragsminimum des betragsmäßig größten Eigenwertes (geeignete Numerierung der η_i und λ_i)

$$\eta_1(\omega) = \frac{\omega^2 \lambda_1^2 - 2(\omega - 1)}{2} + \frac{\omega \lambda_1}{2} \sqrt{\omega^2 \lambda_1^2 - 4(\omega - 1)}$$

von $L(\omega)$ auf. Die Diskussion in Abb. 1 zeigt, daß bei diesem (nichtdifferenzierbaren) Minimum der Radikand den Wert 0 annimmt, d.h. man erhält ω_b in Abhängigkeit vom größten Eigenwert λ_1 von A ('strenge property (A)' und reelle Eigenwerte λ_i vorausgesetzt) als

$$\omega_b^2 \lambda_1^2 - 4(\omega_b - 1) = 0$$

bzw.

$$\omega_b = \frac{2}{1 + \sqrt{1 - \lambda_1^2}} \qquad \ldots\ldots\ldots\ldots \quad (2,12).$$

Der Eigenwert λ_1 ist im allgemeinen nicht bekannt. Seiner iterativen Berechnung bei symmetrischen Matrizen A von 'strenger property (A)' im Laufe der Overrelaxation und mit zunächst willkürlich gewähltem Faktor ω (= 1) sind u.a. die §§ 3-5 dieser Arbeit gewidmet. Dabei ist zu bemerken, daß mit A keinesfalls auch die Matrix $L(\omega)$ der Overrelaxation symmetrisch ist.

§ 3. ITERATIVE BESTIMMUNG VON EIGENWERTEN UND KONVERGENZBESCHLEUNIGUNG NACH THIELE, AITKEN, WYNN SOWIE NEUE FORMELN.

Für Vektoren z im n-dimensionalen EUKLIDISCHEN RAUM benutzen wir die Schreibweise

$$z^{-1} = \frac{1}{z} = \frac{z}{|z|^2}$$

und definieren folgende auf eine Folge von Vektoren (oder Skalaren) $z_0, z_1, \ldots, z_m, z_{m+1}, \ldots$ anwendbare Operatoren $\delta_t z_m$ (Vektor oder Skalar) und $Q z_m$ (Skalar):

$$\delta_t z_m = z_{m+1} - t z_m \quad \text{(t Skalar)} \quad \ldots\ldots\ldots \quad (3,1),$$

$$\delta_t z = z - tz = (1-t)z \quad ,$$

$$\delta z_m \equiv \delta_1 z_m = z_{m+1} - z_m$$

und

$$Q z_m = \frac{z_{m+1}}{z_m} \qquad (z_m \neq 0) \ldots\ldots\ldots\ldots \quad (3,2),$$

$$Q z = 1 \quad .$$

Ferner führen wir den auf eine Folge von Skalaren $\varphi_0, \varphi_1, \ldots, \varphi_m, \varphi_{m+1}, \varphi_{m+2}, \ldots$ anwendbaren 'δ^2-Operator' $P\varphi_m$ (Skalar) von A. C. AITKEN [1] ein:

$$P\varphi_m = \frac{\begin{vmatrix} \varphi_{m+2} & \varphi_{m+1} \\ \varphi_{m+1} & \varphi_m \end{vmatrix}}{\delta^2 \varphi_m} \qquad (\delta^2 \varphi_m \neq 0) \ldots\ldots \quad (3,3).$$

Wichtige Eigenschaften dieser Operatoren sind:

$$\delta_{t_a} \delta_{t_b} = \delta_{t_a} \delta_{t_b} \quad (t_a, t_b \text{ Skalare}) \ldots\ldots\ldots \quad (3,4)$$

$$P(\alpha \varphi_m + \beta) = \alpha P\varphi_m + \beta \quad (\alpha, \beta \text{ Skalare}, \alpha \neq 0) \ldots \quad (3,5)$$

Wir betrachten nun die bei der allgemeinen Iteration

$$x_{m+1} = Lx_m + s \quad (m=0,1,2,\ldots) \ldots\ldots \quad (3,6)$$

zur Lösung von

$$x = Lx + s \quad \ldots\ldots\ldots\ldots \quad (3,7)$$

auftretende Folge der Fehlervektoren $\epsilon_m = x - x_m$

$$\epsilon_m = L\epsilon_{m-1} = \ldots = L^m\epsilon_0 \quad \ldots\ldots\ldots \quad (3,8).$$

Da ein beliebiger Vektor ϵ_0 sich stets als Linearkombination der linear unabhängigen Hauptvektoren $h_1,\ldots,h_n$ der $n \times n$ Koeffizientenmatrix L darstellen läßt, gilt für gewisse skalare Konstanten $a_1,\ldots,a_n$

$$\epsilon_0 = \sum_1^n a_\nu h_\nu \quad \ldots\ldots\ldots\ldots \quad (3,9).$$

In speziellen Fällen können einige der a_ν Null sein, aber für die praktische Anwendung erscheint dies unwahrscheinlich und wird auch im folgenden ausgeschlossen, besonders, da die Elemente der Hauptvektoren gewöhnlich irrationale Zahlen sein werden, während die durch L, s und die gewählte Anfangsnäherung x_0 bestimmten Elemente des Fehlers ϵ_0 meist rationale Zahlen sind. Durch geeignete Wahl von x_0 kann stets $a_\nu \neq 0$ $(\nu = 1,\ldots,n)$ erreicht werden.

Aus (3,9) entnimmt man, daß der iterierte Vektor $\epsilon_m = L^m\epsilon_0$ durch eine Linearkombination der $h_1,\ldots,h_n$ dargestellt sein wird, die bezüglich der Eigenwerte $t_1,\ldots,t_n$ der Matrix L Polynomgestalt besitzt.

Wir beschränken uns im folgenden auf M a t r i z e n L m i t n u r l i n e a r e n E l e m e n t a r t e i l e r n, bei denen die Hauptvektoren $h_1,\ldots,h_n$ Eigenvektoren zu den Eigenwerten $t_1,\ldots,t_n$ von L sind und sich demnach ϵ_m mit den Vektoren $c_\nu = a_\nu h_\nu$ $(c_\nu \neq 0;\ \nu = 1,\ldots,n)$ darstellen läßt als

$$\epsilon_m = \sum_1^n c_\nu t_\nu^m \quad \ldots\ldots\ldots\ldots \quad (3,10).$$

Setzt man voraus, daß die Iteration (3,6) konvergiert, so sind die Eigenwerte $t_1,\dots,t_n$ sämtlich vom Betrage < 1. Auf Grund von $\delta x_m = -\delta \epsilon_m$ gilt dann mit Vektoren $d_\nu = (1-t_\nu)c_\nu (d_\nu \neq 0,$ $\nu = 1,\dots,n)$ die Darstellung

$$\delta x_m = \sum_1^n d_\nu t_\nu^m \quad \dots\dots\dots\dots\dots\dots \quad (3,11).$$

Diese einfache Beschreibung der Folgen ϵ_m und δx_m ermöglicht die Aufstellung von Eliminationsformeln für die gesuchten Eigenwerte $t_1,\dots,t_n$ und die Lösung x aus den berechneten Vektoren $x_m, x_{m+1},\dots$. Die Güte dieser Formeln wird durch die Ordnung $o(|t|^m)$ (LANDAU) oder kurz $o|t|^m$ (bei Skalargleichungen) bzw. Ordnung $\vec{o}(|t|^m,\dots,|t|^m)$ oder kurz $\vec{o}|t|^m$ (bei Vektorgleichungen) des Restgliedes bestimmt.

Die Eigenwerte $t_1,\dots,t_n$ sind durch entsprechende Numerierung wie folgt zu ordnen:

$$|t_1| \geq |t_2| \geq \dots \geq |t_n| \quad \dots\dots\dots\dots\dots \quad (3,12).$$

Darüber hinaus muß zur Herleitung der Eliminationsformeln und zur Bestimmung der Ordnung der Restglieder

$$|t_1| > |t_2| > \dots > |t_q| \quad \dots\dots\dots\dots\dots \quad (3,13)$$

vorausgesetzt werden, wobei q der größte in der Ordnung des Restgliedes vorkommende Index der Eigenwerte t_ν ist ($1 \leq q \leq n$).

Mehrfache Eigenwerte werden aus den Eliminationsformeln nur als einfache Eigenwerte gewonnen, Eigenwerte 0 werden nicht berücksichtigt, da in diesen Fällen in (3,10), (3,11) Summanden zusammengefaßt werden können bzw. Summanden 0 fortfallen. In der Numerierung (3,12), (3,13) werden entsprechend

mehrfache Eigenwerte nur einmal bzw. Eigenwerte 0 nicht mitgezählt und die Anzahl n erniedrigt. Zur Bestimmung von mehrfachen Eigenwerten verweisen wir auf A. C. AITKEN [1].

Zunächst geben wir die im folgenden benötigten ersten vier Glieder einer Reihenentwicklung für $(\delta x_m)^{-1}$ an:

$$(\delta x_m)^{-1} = \frac{d_1 t_1^m + d_2 t_2^m + d_3 t_3^m + \vec{o}|\, t_4|^m}{d_1^2 (t_1^2)^m + 2d_1 d_2 (t_1 t_2)^m + 2d_1 d_3 (t_1 t_3)^m + d_2^2 (t_2^2)^m + o|\, t_2 t_3|^m + o|\, t_1 t_4|^m}$$

$$= \frac{d_1}{d_1^2} \left(\frac{1}{t_1}\right)^m + \left[\frac{d_2}{d_1^2} - 2\frac{d_1}{d_1^2} \cdot \frac{d_1 d_2}{d_1^2}\right] \left(\frac{t_2}{t_1}\right)^m$$

$$+ \left[\frac{d_1}{d_1^2} \left(\frac{d_1 d_2}{d_1^2}\right)^2 - \frac{d_1}{d_1^2} \frac{d_2^2}{d_1^2} - 2\frac{d_1 d_2}{d_1^2}\right] \left(\frac{t_2^2}{t_1^3}\right)^m$$

$$+ \left[\frac{d_3}{d_1^2} - 2\frac{d_1}{d_1^2}\frac{d_1 d_3}{d_1^2}\right] \left(\frac{t_3}{t_1}\right)^m + \vec{o}|\frac{t_2^3}{t_1^4}|^m + \vec{o}|\frac{t_2 t_3}{t_1^3}|^m + \vec{o}|\frac{t_4}{t_1^2}|^m \quad . \quad . \quad (3,14).$$

Daraus gewinnt man

$$Q\,\delta x_m = \frac{\delta x_{m+1}}{\delta x_m}$$

$$= t_1 + k_1 \left(\frac{t_2}{t_1}\right)^m + k_2 \left(\frac{t_2^2}{t_1^2}\right)^m + k_3 \left(\frac{t_3}{t_1}\right)^m + o|\frac{t_2^3}{t_1^3}|^m + o|\frac{t_2 t_3}{t_1^2}| + o|\frac{t_4}{t_1}|^m$$

$$= t_1 + o|\frac{t_2}{t_1}|^m \quad (k_1 \neq 0, \text{ sonst höhere Ordnung}) \quad . \quad . \quad . \quad . \quad (3,15)$$

mit

$$k_1 = -\frac{d_1 d_2}{d_1^2}(t_1 - t_2)$$

$$k_2 = \left[2\left(\frac{d_1 d_2}{d_1^2}\right)^2 - \frac{d_2^2}{d_1^2}\right](t_1 - t_2)$$

$$k_3 = -\frac{d_1 d_3}{d_1^2}(t_1 - t_3) \quad .$$

Eine Formel höherer Ordnung erhält man durch Anwendung des

'δ^2 Operators' von AITKEN (3, 3) auf (3, 15) unter Benutzung der speziellen Eigenschaft (3, 5):

$$PQ\delta x_m = \frac{Q\delta x_{m+2} Q\delta x_m - (Q\delta x_{m+1})^2}{Q\delta x_{m+2} + Q\delta x_m - 2Q\delta x_{m+1}} = t_1 +$$

$$\frac{k_1 k_2 (\frac{t_2^3}{t_1^3})^m (\frac{t_2}{t_1})^2 (\frac{t_1 - t_2}{t_1})^2 + k_1 k_3 (\frac{t_2 t_3}{t_1^2})^m (\frac{t_2 - t_3}{t_1})^2 + o|\frac{t_2^4}{t_1^4}|^m + o|\frac{t_2^2 t_3}{t_1^3}|^m + o|\frac{t_2^2 t_4}{t_1^2}|^m}{k_1 (\frac{t_2}{t_1})^m (\frac{t_1 - t_2}{t_1})^2 + o|\frac{t_2^2}{t_1^2}| + o|\frac{t_3}{t_1}|}$$

$$= t_1 + k_2 (\frac{t_2^2}{t_1^2})^{m+1} + k_3 (\frac{t_3}{t_1})^m (\frac{t_2 - t_3}{t_1 - t_2})^2 + o|\frac{t_2^3}{t_1^3}|^m + o|\frac{t_3^2}{t_1 t_2}| + o|\frac{t_4}{t_1}|^m$$

$$= t_1 + o|\frac{t_2^2}{t_1^2}|^m + o|\frac{t_3}{t_1}|^m \quad \text{. . . . . (3, 16),}$$

$(k_1, k_2, k_3 \neq 0$, sonst höhere Ordnung)

Anwendung des Operators P auf $Q\delta x_m$ hat bewirkt, daß im Restglied von (3, 15) der höchste Koeffizient k_1 verschwand.

Anwendung des Operators P auf $PQ\delta x_m$ bewirkt demnach, daß im Restglied von (3, 16) die größere der beiden Ordnungen $o|\frac{t_2^2}{t_1^2}|^m$, $o|\frac{t_3}{t_1}|^m$ verschwindet.

Mehrfache Anwendung des Operators P ergibt immer bessere Näherungen für t_1.

Ist der erste Eigenwert t_1 hinreichend genau bestimmt, so kann man die Formeln (3, 15) bzw. (3, 16) zur Bestimmung des zweiten Eigenwertes t_2 benutzen, indem man sie nicht auf x_m, sondern auf $\delta_{t_1} x_m$ anwendet. In der Darstellung (3, 11) verschwindet nämlich beim Übergang von δx_m auf $\delta\delta_{t_1} x_m = \delta_{t_1} \delta x_m$ (3, 4) das erste Glied mit t_1^m, und es gilt mit Vektoren

$\tilde{d}_\nu = (t_\nu - t_1)d_\nu$ ($\tilde{d}_\nu \neq 0$, $\nu = 2,\ldots,n$) die Darstellung

$$\delta\delta_{t_1} x_m = \sum_2^n \tilde{d}_\nu t_\nu^m \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (3,17).$$

Das ist bis auf die Numerierung der Indizes (um 1 zu erniedrigen und n auf n' = n-1 herabzusetzen) die gleiche Form wie (3,11). Mit Hilfe dieses Kunstgriffes - der, wie unten ausgeführt, auch mehrfach angewandt werden kann - gewinnt man aus (3,15) bzw. (3,16) neue Formeln zur Bestimmung der höheren Eigenwerte aus den bereits berechneten.

Formeln für höhere Eigenwerte aus (3,15) abgeleitet:

$$\left.\begin{aligned} Q\delta\delta_{t_1} x_m &= \underline{t_2 + o\left|\frac{t_3}{t_2}\right|^m} \\ Q\delta\delta_{t_2}\delta_{t_1} x_m &= \underline{t_3 + o\left|\frac{t_4}{t_3}\right|^m} \\ &\vdots \\ Q\delta\delta_{t_{n-2}}\ldots\delta_{t_1} x_m &= \underline{t_{n-1} + o\left|\frac{t_n}{t_{n-1}}\right|^m} \\ Q\delta\delta_{t_{n-1}}\ldots\delta_{t_1} x_m &= \underline{t_n} \end{aligned}\right\} \quad \ldots\ldots \quad (3,18).$$

Die letzte Formel gibt den exakten Wert für t_n an, gewonnen aus $t_1,\ldots,t_{n-1}$ und $x_m,\ldots,x_{m+n+1}$.

Als Beispiel n = 2:

$$\delta x_m = x_{m+1} - x_m = d_1 t_1^m + d_2 t_2^m$$

$$\delta_{t_1} x_m = x_{m+1} - t_1 x_m = (t_2 - t_1) d_2 t_2^m = \tilde{d}_2 t_2^m$$

$$Q\delta\delta_{t_1} x_m = \frac{(x_{m+3} - t_1 x_{m+2}) - (x_{m+2} - t_1 x_{m+1})}{(x_{m+2} - t_1 x_{m+1}) - (x_{m+1} - t_1 x_m)}$$

$$= \frac{\tilde{d}_2 t_2^{m+2} - \tilde{d}_2 t_2^{m+1}}{\tilde{d}_2 t_2^{m+1} - \tilde{d}_2 t_2^m} = t_2$$

Formeln für höhere Eigenwerte aus (3,16) abgeleitet:

$$\left.\begin{aligned} PQ\delta\delta_{t_1} x_m &= t_2 + o\left|\frac{t_3^2}{t_2^2}\right|^m + o\left|\frac{t_4}{t_2}\right|^m \\ PQ\delta\delta_{t_2}\delta_{t_1} x_m &= t_3 + o\left|\frac{t_4^2}{t_3^2}\right|^m + o\left|\frac{t_5}{t_3}\right|^m \\ &\vdots \\ PQ\delta\delta_{t_{n-3}}\cdots\delta_{t_1} x_m &= t_{n-2} + o\left|\frac{t_{n-1}^2}{t_{n-2}^2}\right|^m + o\left|\frac{t_n}{t_{n-2}}\right|^m \\ PQ\delta\delta_{t_{n-2}}\cdots\delta_{t_1} x_m &= t_{n-1} + o\left|\frac{t_n^2}{t_{n-1}^2}\right|^m \end{aligned}\right\} \quad . . \ (3,19).$$

Eine (3,18) entsprechende letzte Formel zur exakten Bestimmung von t_n fehlt, da P nicht auf $Q\delta\delta_{t_{n-1}}\cdots\delta_{t_1} x_m = t_n$ angewandt werden darf (3,3).

Die Ordnung der Formeln (3,19) läßt sich wie bei (3,16) durch mehrfache Anwendung des Operators P weiter erhöhen.

A. C. AITKEN [1] empfiehlt zur Berechnung des zweiten Eigenwertes aus dem ersten folgende Formel (wir geben ohne Beweis die berechnete Ordnung an):

$$t_1 \cdot Q\delta Q\delta x_m = t_2 + o\left|\frac{t_3}{t_2}\right|^m + o\left|\frac{t_2}{t_1}\right|^m \quad . . . . . \ (3,20).$$

Da $\left|\frac{t_2}{t_1}\right| > \left|\frac{t_3}{t_2}\right|$ ausfallen kann, ist die vergleichbare neue Formel (3,18) (in deren Restglied t_1 überhaupt nicht auftritt) besser als diese Formel von A. C. AITKEN.

Konvergenzbeschleunigung.

Wie die Darstellung (3,10) zeigt, liefert das Iterationsverfahren (3,6) Näherungswerte x_m, welche die exakte Lösung x von (3,7)

bis auf ein Restglied der Ordnung $\vec{o}|\,t_1|^m$ wiedergeben:

$$x_m = x - \sum_1^n c_\nu t_\nu^m \qquad (m = 0, 1, 2, \ldots)$$

$$= \underline{x + \vec{o}|\,t_1|^m} \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (3,21).$$

Will man die Konvergenz beschleunigen, d.h. den Wert x mit höherer Ordnung als $|\,t_1|^m$ approximieren, so kann man ähnlich wie bei der Berechnung der Eigenwerte aus (3,21) Eliminationsformeln höherer Ordnung ableiten.

Auf einem Kolloquium in Hamburg 1961 hat Prof. H. RUTISHAUSER folgende bisher unveröffentlichte Formel von P. WYNN mitgeteilt:

$$x \sim x_{m+1} + \frac{1}{\dfrac{1}{x_{m+2}-x_{m+1}} - \dfrac{1}{x_{m+1}-x_m}} \qquad \ldots\ldots \quad (3,22).$$

$$(m = 0, 1, 2, \ldots)$$

Wir formulieren (3,22) in der bisher benutzten Operatorschreibweise und bestimmen unter Benutzung der Reihenentwicklung (3,14) die Ordnung des Restgliedes:

$$x_{m+1} + (\delta(\delta x_m)^{-1})^{-1} = x_{m+1} +$$

$$\frac{\frac{d_1}{2}(\frac{1}{t_1})^m \frac{1-t_1}{t_1} + (\frac{d_2}{2} - 2\frac{d_1}{2}\frac{d_1 d_2}{2})(\frac{t_2}{2})^m \frac{t_2^2 - t_1^2}{t_1^2} + \vec{o}|\,\frac{t_2^2}{t_1^3}|^m + \vec{o}|\,\frac{t_3}{t_1^2}|^m}{\frac{1}{d_1^2}(\frac{1}{t_1})^m (\frac{1-t_1}{t_1})^2 - 2\frac{1}{d_1^2}\frac{d_1 d_2}{d_1^2}(\frac{t_2}{t_1})^m \frac{(t_2-t_1^2)(1-t_1)}{t_1^3} + o|\,\frac{t_2^2}{t_1^3}|^m + o|\,\frac{t_3}{t_1^2}|^m}$$

$$= x_{m+1} + (1-t_1)^{-2}\left[d_1 t_1^{m+1}(1-t_1) + d_2 t_2^m (t_2 - t_1^2) + \vec{o}|\,\frac{t_2^2}{t_1}|^m + \vec{o}|\,t_3|^m\right]$$

$$= x - c_2 t_2^m (t_1 - t_2)^2 (1-t_1)^{-2} + \vec{o}|\,\frac{t_2^2}{t_1}|^m + \vec{o}|\,t_3|^m$$

$$= \underline{x + \vec{o}|\,t_2|^m} \qquad \ldots\ldots\ldots\ldots\ldots\ldots \quad (3,23).$$

Von THIELE (Interpolationsrechnung 1909), der erstmalig Formeln zur Konvergenzbeschleunigung aufstellte, und später unabhängig von A. C. AITKEN [1] wiederentdeckt, stammt folgende Formel, die von beiden Autoren aber nicht wie hier auf Vektoren x_m, x , sondern auf deren Komponenten $x_m^{(i)}, x^{(i)}$ (i=1,...,n) angewandt wurde:

$$x_{m+2} - (\delta x_{m+1})^2 \cdot (\delta^2 x_m)^{-1} = x_{m+2} -$$

$$\left[d_1^2 (t_1^2)^{m+1} + 2 d_1 d_2 (t_1 t_2)^{m+1} + o|\, t_1 t_3 |^m + o|\, t_2^2 |^m \right] \cdot$$

$$\cdot \left[\frac{d_1^*}{d_1^{*2}} \left(\frac{1}{t_1}\right)^m + \left[\frac{d_2^*}{d_1^{*2}} - 2 \frac{d_1^*}{d_1^{*2}} \frac{d_1^* d_2^*}{d_1^{*2}} \right] \left(\frac{t_2}{t_1}\right)^m + \vec{o}|\, \frac{t_2^2}{t_1^3} |^m + \vec{o}|\, \frac{t_3}{t_1^2} |^m \right] =$$

(Zwischenrechnung ausgelassen)

$$= \underline{x + \vec{o}|\, t_2 |^m} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (3,24).$$

Mit der Abkürzung $d_\nu^* = (t_\nu - 1) d_\nu$ kann man zur Herleitung von (3,24) wiederum die Reihenentwicklung (3,14) benutzen.

Diese Formel von THIELE, AITKEN hat wie die Formel von P. WYNN ein Restglied der Ordnung $\vec{o}|\, t_2 |^m$; sie ist bei Übergang von Vektoren x_m auf Komponenten $x_m^{(i)}$ (i=1,...,n) mit der Formel von P. WYNN identisch:

$$x_{m+2}^{(i)} - (\delta x_{m+1}^{(i)})^2 (\delta^2 x_m^{(i)})^{-1} =$$

$$= x_{m+2}^{(i)} - \frac{x_{m+2}^{(i)\,2} - 2 x_{m+2}^{(i)} + x_{m+1}^{(i)}}{x_{m+2}^{(i)} - 2 x_{m+1}^{(i)} + x_m^{(i)}}$$

$$= x_{m+1}^{(i)} + \frac{(x_{m+2}^{(i)} - x_{m+1}^{(i)})(-x_{m+1}^{(i)} + x_m^{(i)})}{x_{m+2}^{(i)} - 2 x_{m+1}^{(i)} + x_m^{(i)}}$$

$$= x_{m+1}^{(i)} + (\delta(\delta x_m^{(i)})^{-1})^{-1} \quad ,$$

stimmt aber in Vektorschreibweise nicht mit (3,23) überein, wie das Gegenbeispiel $x_m = (1,0)$, $x_{m+1} = (0,0)$, $x_{m+2} = (0,1)$ zeigt:

$$x_{m+2} - (\delta x_{m+1})^2(\delta^2 x_m)^{-1} = (0,1) - (0,1)^2(1,1)^{-1}$$
$$= (-\tfrac{1}{2}, \tfrac{1}{2})$$
$$x_{m+1} + (\delta(\delta x_m)^{-1})^{-1} = (0,0) + \left[(0,1)^{-1} - (-1,0)^{-1}\right]^{-1}$$
$$= (\tfrac{1}{2}, \tfrac{1}{2}) \quad .$$

Im folgenden wird die Formel (3,23) von P. WYNN bevorzugt, da sie sich zur Programmierung auf elektronischen Rechenanlagen besser eignet als die Formel (3,24) von THIELE, AITKEN.

Die Vektorschreibweise der Formeln (3,23), (3,24) hat gegenüber komponentenweiser Konvergenzbeschleunigung den Vorteil, daß in der Formel 'im Nenner' auftretende Vektoren von Null verschieden sind, solange die exakte Lösung x noch nicht erreicht worden ist. Bei der Programmierung muß aber darauf geachtet werden, daß die Konvergenzbeschleunigung aussetzt, wenn die Differenzen δx_m zu klein werden und Rundungsfehler überwiegen.

Wir wenden nun, wie bei der Bestimmung höherer Eigenwerte, den Kunstgriff an, x und x_m durch $\delta_{t_1} x$ und $\delta_{t_1} x_m$ zu ersetzen. Dann erhöhen sich die Indizes der die Restglieder bestimmenden Eigenwerte jeweils um 1, und wir erhalten neue Formeln höherer Ordnung. (Übergang von d_ν bzw. c_ν auf $\tilde{d}_\nu = (t_\nu - t_1)d_\nu$ bzw. $\tilde{c}_\nu = (t_\nu - t_1)c_\nu$, $\nu = 1,\ldots,n$). Die zur Bildung von $\delta_{t_1} x_m$ und entsprechend $\delta_{t_2}\delta_{t_1} x_m$ u.s.w. benötigten Eigen-

werte $t_1, t_2, \ldots$ bestimmt man vorher nach Formel (3,16),(3,19).

Formeln höherer Ordnung aus (3,23) abgeleitet:

Fassung I:

$$\left\{\begin{array}{l}
\delta_{t_1} x_{m+1} + (\delta(\delta\,\delta_{t_1} x_m)^{-1})^{-1} = \underline{\delta_{t_1} x + \vec{o}|\, t_3|^{\,m}} \\
\delta_{t_2}\delta_{t_1} x_{m+1} + (\delta(\delta\,\delta_{t_2}\delta_{t_1} x_m)^{-1})^{-1} = \underline{\delta_{t_2}\delta_{t_1} x + \vec{o}|\, t_4|^{\,m}} \\
\quad\vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \ldots\ldots \qquad (3,25a) \\
\delta_{t_{n-2}} \ldots \delta_{t_1} x_{m+1} + (\delta(\delta\,\delta_{t_{n-2}} \ldots \delta_{t_1} x_m)^{-1})^{-1} = \underline{\delta_{t_{n-2}} \ldots \delta_{t_1} x + \vec{o}|\, t_n|^{\,m}} \\
\delta_{t_{n-1}} \ldots \delta_{t_1} x_{m+1} + (\delta(\delta\,\delta_{t_{n-1}} \ldots \delta_{t_1} x_m)^{-1})^{-1} = \underline{\delta_{t_{n-2}} \ldots \delta_{t_1} x}
\end{array}\right.$$

Fassung II:

$$\left\{\begin{array}{l}
(1-t_1)^{-1}\left[\delta_{t_1} x_{m+1} + (\delta(\delta\,\delta_{t_1} x_m)^{-1})^{-1}\right] = \underline{x+\vec{o}|\, t_3|^{\,m}} \\
(1-t_1)^{-1}(1-t_2)^{-1}\left[\delta_{t_2}\delta_{t_1} x_{m+1} + (\delta(\delta\,\delta_{t_2}\delta_{t_1} x_m)^{-1})^{-1}\right] = \underline{x+\vec{o}|\, t_4|^{\,m}} \\
\quad\vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \ldots\ldots \qquad (3,25) \\
(1-t_1)^{-1} \ldots (1-t_{n-2})^{-1}\left[\delta_{t_{n-2}} \ldots \delta_{t_1} x_{m+1} + (\delta(\delta\,\delta_{t_{n-2}} \ldots \delta_{t_1} x_m)^{-1})^{-1}\right] = \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad = \underline{x+\vec{o}|\, t_n|^{\,m}} \\
(1-t_1)^{-1} \ldots (1-t_{n-1})^{-1}\left[\delta_{t_{n-1}} \ldots \delta_{t_1} x_{m+1} + (\delta(\delta\,\delta_{t_{n-1}} \ldots \delta_{t_1} x_m)^{-1})^{-1}\right] = \underline{x}
\end{array}\right.$$

Die letzte Formel gibt den exakten Wert für x an, gewonnen aus

$t_1, \ldots, t_{n-1}$ und $x_m, \ldots, x_{m+n+1}$.

Als Beispiel n = 2:

$\delta\,\delta_{t_1} x_m = (x_{m+2} - t_1 x_{m+1}) - (x_{m+1} - t_1 x_m) = \tilde{d}_2 t_2^{\,m}$ vgl. (3,17)

$(\delta\delta_{t_1} x_m)^{-1} = \tilde{d}_2\, \tilde{d}_2^{-2} t_2^{-m}$

$$x = x_{m+2} + c_1 t_1^{m+2} + c_2 t_2^{m+2} \qquad \text{vgl. (3,21)}$$

$$t_1 x = t_1 x_{m+1} + c_1 t_1^{m+2} + c_2 t_1 t_2^{m+1}$$

$$(1-t_1)x = \delta_{t_1} x_{m+1} + c_2 (t_2 - t_1) t_2^{m+1}$$

$$\delta_{t_1} x_{m+1} = (1-t_1)x - (1-t_2)^{-1} \tilde{d}_2 t_2^{m+1}$$

$$(1-t_1)^{-1} \left[\delta_{t_1} x_{m+1} + (\delta(\delta\, \delta_{t_1} x_m)^{-1})^{-1} \right] = \frac{1}{1-t_1} \left[(x_{m+2} - t_1 x_{m+1}) + \right.$$

$$\left. + \frac{1}{\frac{1}{(x_{m+3} - t_1 x_{m+2}) - (x_{m+2} - t_1 x_{m+1})} - \frac{1}{(x_{m+2} - t_1 x_{m+1}) - (x_{m+1} - t_1 x_m)}} \right]$$

$$= \frac{1}{1-t_1} \left[(1-t_1)x - (1-t_2)^{-1} \tilde{d}_2 t_2^{m+1} + \frac{\tilde{d}_2 t_2^{m}}{\tilde{d}_2 (t_2^{-1} - 1)} \right] = x \,.$$

§ 4. EIN SATZ ÜBER DIE DARSTELLUNG DER FEHLERVEKTOREN BEI DER OVER-RELAXATION.

Die Formeln des vorigen Paragraphen zur iterativen Bestimmung von Eigenwerten und zur Konvergenzbeschleunigung lassen sich auf die Overrelaxation anwenden, wenn nachgewiesen werden kann, daß die bei der Overrelaxation (2, 2) auftretenden Fehlervektoren $\epsilon_m = x - x_m$ $(m = 0, 1, 2, \ldots)$ sich mit den Eigenvektoren $\eta_1(\omega), \ldots, \eta_n(\omega)$ von $L(\omega)$ (2,4) und Vektoren $c_\nu \neq 0$ $(\nu = 1, \ldots, n)$ darstellen lassen als

$$\epsilon_m = \sum_1^n c_\nu \eta_\nu^m \qquad \text{(vgl. (3,10))}.$$

Wie FORSYTHE, WASOW [10] bemerken, ist in der Literatur bisher kein Beweis bekannt.

Wir bringen einen solchen Beweis für symmetrische Matrizen A 'von strenger property (A)', und zwar in (4, 1) für Matrizen $L(\omega)$ mit $\omega \neq 1$ und in (4, 4) für Matrizen $L(1)$. In (4, 1) wird sogar allgemeiner bewiesen, daß die Matrix $L(\omega)$ $(\omega \neq 1)$ diagonalisierbar ist und daher nur lineare Elementarteiler besitzt. Für Matrizen mit nur linearen Elementarteilern war die Darstellung (3,10) in § 3 hergeleitet worden.

Satz (4, 1): Symmetrie und 'strenge property (A)' (2,6) der Matrix A sind hinreichend dafür, daß die bei der Overrelaxation (2, 2) auftretende Matrix (2,4)

$$L(\omega) = (E - \omega A_L)^{-1}((1-\omega)E + \omega A_R)$$

für $\omega \neq 1$ (komplex) nur lineare Elementarteiler besitzt.

Beweis: Aus der durch 'strenge property (A)' bedingten speziellen Rechtecksgestalt der Matrix A_L folgt unmittelbar

$$A_L^2 = 0 \quad , \qquad \text{daraus}$$

$$(E-\omega A_L)^{-1} = E + \omega A_L$$

und $$L(\omega) = (E + \omega A_L)((1-\omega)E + \omega A_R)$$

$$= \omega(1-\omega)A_L + (1-\omega)E + \omega^2 A_L A_R + \omega A_R \quad . \; . \; . \quad (4,2).$$

Für $\omega \neq 1$ findet man unter Berücksichtigung von $A_L = A_R'$ (Symmetrie von A) eine nicht singuläre Matrix M so, daß $M^{-1}L(\omega)M$ (komplex) symmetrisch ist:

$$M^{-1} \qquad\qquad L(\omega) \qquad\qquad M$$

$$\left(\begin{array}{c|c} (1-\omega)^{\frac{1}{2}}E_1 & 0 \\ \hline 0 & E_2 \end{array}\right)\left(\begin{array}{c|c} (1-\omega)E_1 & \omega A_R \\ \hline \omega(1-\omega)A_L & (1-\omega)E_2 + \omega^2 A_L A_R \end{array}\right)\left(\begin{array}{c|c} (1-\omega)^{-\frac{1}{2}}E_1 & 0 \\ \hline 0 & E_2 \end{array}\right) =$$

$$= \left(\begin{array}{c|c} (1-\omega)E_1 & \omega(1-\omega)^{\frac{1}{2}}A_R \\ \hline \omega(1-\omega)^{\frac{1}{2}}A_L & (1-\omega)E_2 + \omega^2 A_L A_R \end{array}\right) \quad . \; . \; . \; . \; . \; . \quad (4,3).$$

$$M^{-1}L(\omega)M$$

E_1 bzw. E_2 sind Einheitsmatrizen der Ordnung n_1 bzw. n_2.

Die (komplex) symmetrische Matrix $M^{-1}L(\omega)M$ ist der Matrix $L(\omega)$ äquivalent und $L(\omega)$ daher diagonalisierbar (q.e.d.).

Satz (4,4): Symmetrie und 'strenge property (A)' (2,6) sind hinreichend dafür, daß die beim Einzelschrittverfahren

- Overrelaxation (2,3) im Spezialfall $\omega = 1$ mit der Matrix (2,4)

$$L(1) = (E - A_L)^{-1} A_R \quad -$$

auftretenden Fehlervektoren $\epsilon_m = x - x_m$ $(m = 0, 1, 2, \ldots)$ sich mit Hilfe der Eigenwerte $\eta_1(1), \ldots, \eta_n(1)$ von L(1) und Vektoren $c_\nu \neq 0$ $(\nu = 1, \ldots, n)$ darstellen lassen als

$$\epsilon_m = \sum_1^n c_\nu \; \overset{m}{\eta_\nu}(1) \quad .$$

B e w e i s : Aus dem Beweis des vorigen Satzes entnehmen wir

$$L(1) = (E+A_L)A_R = \begin{array}{c} \\ n_1 \\ n_2 \end{array} \overset{\overset{n_1}{\longleftrightarrow} \quad \overset{n_2}{\longleftrightarrow}}{\begin{pmatrix} 0 & A_R \\ 0 & A_L A_R \end{pmatrix}} \qquad \ldots\ldots\ldots (4,5).$$

Entsprechend der Anzahl Zeilen n_1 und Spalten n_2 von A_R und Zeilen n_2 und Spalten n_2 von $A_L A_R$ berechnen wir den Vektor $\epsilon_m^{(1)}$ der ersten n_1 Komponente und den Vektor $\epsilon_m^{(2)}$ der übrigen n_2 Komponenten von ϵ_m getrennt:

$$\begin{aligned} \epsilon_m^{(1)} &= A_R \epsilon_{m-1}^{(2)} \\ \epsilon_m^{(2)} &= A_L A_R \epsilon_{m-1}^{(2)} = \ldots = (A_L A_R)^m \epsilon_0^{(2)} \end{aligned} \qquad \ldots\ldots\ldots (4,6).$$

Da mit A auch $A_L A_R$ symmetrisch ist, kann man $\epsilon_0^{(2)}$ durch eine Linearkombination der n_2 linear unabhängigen Eigenvektoren f_ν $(\nu = 1, \ldots, n_2)$ von $A_L A_R$ mit Konstanten b_ν ($b_\nu \neq 0$, $\nu = 1, \ldots, n_2$ Vgl. (3,9)) darstellen:

$$\epsilon_0^{(2)} = \sum_1^{n_2} b_\nu f_\nu \qquad \ldots\ldots\ldots\ldots (4,7).$$

Bei der speziellen Gestalt (4,5) von L(1) gilt

$$\det(L(1) - \eta E) = (-1)^{n_1} \eta^{n_1} \det(A_L A_R - \eta E_2)$$

d.h. die Eigenwerte von $A_L A_R$ sind Eigenwerte $\eta_1, \ldots, \eta_{n_2}$ von L(1) und unter diesen befinden sich alle von Null verschiedenen Eigenwerte von L(1).

Wir erhalten mit Vektoren $g_\nu = b_\nu f_\nu$ $(g_\nu \neq 0,\ \nu = 1,\dots,n_2)$

$$\epsilon_m^{(2)} = \sum_1^{n_2} g_\nu \eta_\nu^m \ ;$$

$$\epsilon_m^{(1)} = \sum_1^{n_2} A_R g_\nu \eta_\nu^{m-1}$$

und daraus mit Vektoren $c_\nu \neq 0$:

$$\left.\begin{cases} c_\nu^{(1)} = \begin{cases} A_R g_\nu \eta_\nu^{-1} & , \text{ falls } \eta_\nu \neq 0 \\ \text{beliebig} & , \text{ falls } \eta_\nu = 0 \end{cases} \\ c_\nu^{(2)} = g_\nu \end{cases} \right\} c_\nu \quad (\nu = 1,\dots,n_2) \ ,$$

$c_\nu \neq 0$ beliebig für $\nu = n_2+1,\dots,n$, die gesuchte Darstellung

$$\epsilon_m = \sum_1^n c_\nu \eta_\nu^m \qquad \text{(q.e.d.)} .$$

§5. ITERATIVE BESTIMMUNG DES OPTIMALEN RELAXATIONSFAKTORS ω_b UND KONVERGENZBESCHLEUNIGUNG BEI DER OVERRELAXATION

Nach den Sätzen von § 4 ist bei der Overrelaxation mit symmetrischer Matrix A 'von strenger property (A)' die Anwendung der Formeln von § 3 zur iterativen Bestimmung von Eigenwerten und zur Konvergenzbeschleunigung möglich. Die iterative Bestimmung von Eigenwerten ist eine wichtige Ergänzung zur YOUNGschen Theorie (§ 2), nach der ω_b aus dem betragsmäßig größten Eigenwert von A berechnet werden kann.

Um die in den Restgliedern der Formeln bei Anwendung auf die Overrelaxation entstehenden Ordnungen näher bestimmen zu können, werden in Abb. 2 der Verlauf von

$$\left|\frac{\eta_2(\omega)}{\eta_1(\omega)}\right| \quad , \quad \left|\frac{\eta_2^2(\omega)}{\eta_1^2(\omega)}\right| \quad , \quad \left|\frac{\eta_3(\omega)}{\eta_1(\omega)}\right| \quad , \quad \left|\frac{\eta_3(\omega)}{\eta_2(\omega)}\right|$$

für spezielle $\lambda_1 = 0{,}98$, $\lambda_2 = 0{,}9$, $\lambda_3 = 0{,}8$ angegeben und die Extremwerte dieser Funktionen allgemein für reelle (da A symmetrisch) $\lambda_i < 1$ und reelle ω aus

$$\eta_i(\omega) = \frac{\omega^2\lambda_i^2 - 2(\omega-1)}{2} + \frac{\omega\lambda_i}{2}\sqrt{\omega^2\lambda_i^2 - 4(\omega-1)} \ldots \text{ vgl. (2,11)}$$

berechnet (vgl. auch Abb. 1).

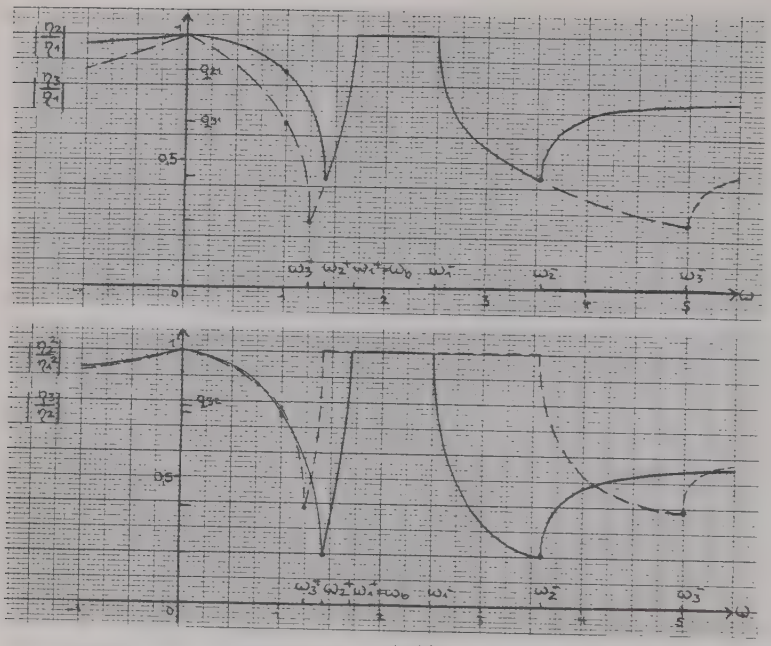


Abb. 2 Darstellung von $\left|\frac{\eta_2(\omega)}{\eta_1(\omega)}\right|$ u.a. für spezielle $\lambda_1, \lambda_2, \ldots$

Es ist $\omega_i^{\pm} = \dfrac{2}{1 \pm \sqrt{1-\lambda_i^2}}$, $\omega_b = \omega_1^+$

und $\left|\dfrac{\eta_i(\omega)}{\eta_j(\omega)}\right| \equiv q_{ij}(\omega) = q_{ij}\left(\dfrac{\omega}{\omega-1}\right)$,

$$q_{ij}(1) = q_{ij} = \left|\frac{\lambda_i}{\lambda_j}\right|^2$$

$$q_{ij}(\omega_i^{\pm}) = \frac{q_{ij}}{\left(1+\sqrt{1-q_{ij}}\right)^2}$$

$$q_{ij}(\omega_j^{\pm}) = 1 \qquad (i,j=1,\ldots,n)$$

. . . . . (5,1)

Aus (5,1) und Abb. 2 ist zu entnehmen, daß die iterative Bestimmung z.B. des ersten Eigenwertes $\eta_1(\omega)$ der Overrelaxationsmatrix $L(\omega)$ (etwa) nach dem δ^2-Verfahren von A.C.AITKEN

$$PQ\delta x_m = \eta_1(\omega) + o\left|\frac{\eta_2(\omega)^2}{\eta_1(\omega)^2}\right|^m + o\left|\frac{\eta_3(\omega)}{\eta_1(\omega)}\right|^m \quad \text{... vgl. (3,16)}$$

$$(\eta_1(\omega) > \eta_2(\omega) > \eta_3(\omega) \geq \ldots \geq |\eta_n(\omega)|)$$

auch für $\omega < 0$ oder $\omega > \omega_1^- > 2$ möglich ist, obwohl für diese Werte von ω die Folge der x_m nach Satz (2,11a) divergiert!

Wird der Eigenwert $\eta_1(\omega)$ aber nur zur Bestimmung des optimalen Relaxationsfaktors ω_b berechnet, so empfiehlt es sich, im Bereich $0 < \omega < \omega_1^+ < 2$ zu iterieren, um Ausgangsnäherungen x_m für die spätere Overrelaxation mit ω_b zu erhalten.

Konvergenzbeschleunigung, etwa nach der neuen Formel

$$\frac{\delta\,\eta_1(\omega)^{x_m} + (\delta(\delta\,\eta_1(\omega)^{x_m})^{-1})^{-1}}{1 - \eta_1(\omega)} = x + \vec{\delta}\,|\eta_3(\omega)|^m \quad \text{vgl. (3,25)}$$

$$(1 > |\eta_1(\omega)| > |\eta_2(\omega)| > |\eta_3(\omega)| \geq \ldots \geq |\eta_n(\omega)|)$$

ist bei der Overrelaxation nur dann sinnvoll, wenn der optimale Relaxationsfaktor ω_b noch nicht (iterativ) berechnet worden ist. Um die Formeln zur Konvergenzbeschleunigung anwenden zu können, muß nämlich wegen Voraussetzung $|\eta_1(\omega)| > |\eta_2(\omega)|$ mit einem Relaxationsfaktor $\omega < \omega_1^+ = \omega_b$ iteriert werden (5, 1). Das hat - in Abb. 1 erkennt man die senkrechte Tangente am linken Teil der Kurve $\eta_1(\omega)$ bei ω_b - ein starkes Anwachsen von $\eta_1(\omega)$ und damit eine sehr viel langsamere Konvergenz der Overrelaxation zur Folge (insbesondere bei großen Gleichungssystemen mit i.a. großen Eigenwerten λ_1, vgl. Abb. 1).

(5,2) Eine Iterative Methode zur Lösung linearer Gleichungssysteme x = Ax + r mit symmetrischer Matrix A 'von strenger property (A)'.

a) Ausgehend von einem beliebigen Vektor x_0 werden Näherungen x_{m+1} $(m = 0, 1, 2, \ldots)$ nach dem Einzelschrittverfahren (GAUSS [12], SEIDEL [21]; $\omega = 1$) berechnet:

$$x_{m+1} = A_L x_{m+1} + A_R x_m + r \quad \ldots\ldots \text{ vgl. (2,2)},$$

und es wird der erste Eigenwert $\eta_1(1)$ der Matrix $L(1) = (E - A_L)^{-1} A_R$ nach dem δ^2-Verfahren von A.C.AITKEN [1] bestimmt:

$$PQ\delta x_m = \eta_1(1) + o\left|\frac{\eta_2(1)^2}{\eta_1(1)^2}\right|^m + o\left|\frac{\eta_3(1)}{\eta_1(1)}\right|^m \quad \ldots \text{ vgl. (3,16)}$$

$$(|\eta_1(1)| > |\eta_2(1)| > |\eta_3(1)| \geq \ldots \geq |\eta_n(1)|)$$

Schema zu a):

x_m			
	δx_m		
x_{m+1}		$Q\delta x_m$	
	δx_{m+1}		
x_{m+2}		$Q\delta x_{m+1}$	$PQ\delta x_m \sim \eta_1(1)$
	δx_{m+2}		
x_{m+3}		$Q\delta x_{m+2}$	$PQ\delta x_{m+1} \sim \eta_1(1)$
	δx_{m+3}		
x_{m+4}		$Q\delta x_{m+3}$	. .
	δx_{m+4}	.	
x_{m+5}			
.	.		

Aus $\eta_1(1)$ wird nach der Theorie von D. YOUNG [26]

— vgl. (2,11), (2,12), $\eta_1(1) = \lambda_1^2$ — der optimale Overrelaxationsfaktor ω_b bestimmt:

$$\omega_b = \frac{2}{1+\sqrt{1-\eta_1(1)}} \quad \text{. . . . . vgl. (2,12) .}$$

Die Iteration a) wird beendet, wenn sich (etwa) die dritte Stelle von ω_b nicht mehr ändert. Diese ist nach oben aufzurunden.

b) Auf die letzten 4 Werte x_m der Iteration a) wird mit dem für $\eta_1(1)$ gewonnenen Wert die neue Formel höherer Ordnung (3,25) zur Konvergenzbeschleunigung angewandt :

$$\frac{\delta\,\eta_1(1)^{x_{m+1}} + (\delta(\delta\,\delta\,\eta_1(1)^{x_m})^{-1})^{-1}}{1-\eta_1(1)} = x + \vec{o}\,|\,\eta_3(1)|^m \quad \text{. . . vgl. (3,25)}$$

$$(1>|\eta_1(1)|>|\eta_2(1)|>|\eta_3(1)|\geq\ldots\geq|\eta_n(1)|) .$$

Schema zu b):

x_m			
	$\delta\,\eta_1(1)^{x_m}$		
x_{m+1}		$\delta\,\delta\,\eta_1(1)^{x_m}$	
	$\delta\,\eta_1(1)^{x_{m+1}}$		$(\delta(\delta\,\delta\,\eta_1(1)^{x_m})^{-1})^{-1}$
x_{m+2}		$\delta\,\delta\,\eta_1(1)^{x_{m+1}}$	
	$\delta\,\eta_1(1)^{x_{m+1}}$		
x_{m+3}			

c) Ausgehend von dem in b) für x berechneten neuen Wert x_0 werden Näherungen x_{m+1} $(m = 0, 1, 2, \ldots)$ nach der Overrelaxation von SOUTHWELL [22], YOUNG [26] berechnet (ω_b aus a)):

$$x_{m+1} = x_m + \omega_b(A_L x_{m+1} - E x_m + A_R x_m + r) \quad \text{. . . . vgl. (2,2) .}$$

Diese Iterative Methode erfordert zu ihrer Anwendung keine vorherige Kenntnis von Eigenwerten. Die langsame Konvergenz des im ersten Teil a) benutzten Einzelschrittverfahrens wird

weitgehend ausgeglichen durch die im Teil b) benutzte Konvergenzbeschleunigung höherer Ordnung, so daß die Iterative Methode a), b), c) so schnell konvergiert wie Overrelaxation mit als bekannt vorausgesetzten optimalem Faktor ω_b.

Die benutzten Formeln (3,16) und (3,25) sind auf Grund des Satzes (4,4) anwendbar und die Voraussetzungen

$$|\eta_1(1)| \geq |\eta_2(1)| \geq |\eta_3(1)|$$

entweder erfüllt oder - wegen 'strenger property (A)' sind alle Eigenwerte $\eta_i(1)$ reell nicht-negativ (2,11) und aus $|\eta_i(1)| = |\eta_j(1)|$ folgt $\eta_i(1) = \eta_j(1)$ - es treten mehrfache Eigenwerte auf, die in den Formeln nur als einfache Eigenwerte gezählt werden (vgl. 3,13), was durch Fortfallen von $\eta_3(1)$ oder $\eta_2(1)$, $\eta_3(1)$, d.h. durch Fortfall von Restgliedern zum Ausdruck kommt. Die Voraussetzung $1 > |\eta_1(1)|$ ist dann erfüllt, wenn das Einzelschrittverfahren konvergiert.

Satz (5,3): Die Iterative Methode (5,3) konvergiert genau dann gegen die dann eindeutig existierende Lösung von $x = Ax + r$, A symmetrisch und 'von strenger property (A)' vorausgesetzt, wenn das Einzelschrittverfahren a) konvergiert.

Der Beweis folgt unmittelbar aus Satz (2,11a), wenn man berücksichtigt, daß auf Grund der Symmetrie von A alle Eigenwerte λ_i reell sind und der aus (2,12) berechnete Wert ω_b reell ist, wenn $\lambda_1^2 < 1$ ist, d.h. wenn das Einzelschrittverfahren konvergiert.

Einfache Konvergenzkriterien zum Einzelschrittverfahren und damit zu dieser Iterativen Methode werden im § 6 hergeleitet.

§ 6. FEHLERABSCHÄTZUNGEN ZUM EINZELSCHRITTVERFAHREN.

Zum Einzelschrittverfahren ($\omega = 1$) sind von L. COLLATZ [5], W. DÜCK [7], H. SASSENFELD [19], J. SCHRÖDER [20] u.a.m. numerisch brauchbare Fehlerabschätzungen veröffentlicht worden. In diesem Kapitel werden dazu zwei neue Fehlerabschätzungen für beliebige Matrixnormen hergeleitet, von denen die erste [9], gültig für beliebige Matrizen A, die Dreiecksform der Matrizen A_L und A_R und die zweite, gültig für Matrizen A 'von strenger property (A)', die spezielle Rechtecksform der Matrizen A_L und A_R benutzt. Außerdem werden zwei neue hinreichende Konvergenzkriterien sowie allgemeiner als bisher die Fehlerabschätzung von W. DÜCK für beliebige und die Fehlerabschätzung von L. COLLATZ für monotone, absolute Matrixnormen hergeleitet. Bei der Aufstellung einer numerisch brauchbaren Fehlerabschätzung zur Overrelaxation im Falle $\omega \neq 1$ tritt die Schwierigkeit auf, die Spektralnorm der i.a. nicht hermiteschen Matrix $L(\omega)$ mit erträglichem Rechenaufwand zu bestimmen. Wir verweisen auf die Fehlerabschätzung von J. ALBRECHT [3].

1. Hilfsmittel.

Der Raum R_n der Vektoren x und der Raum M_n der quadratischen Matrizen A sind nach Einführung einer Vektor- bzw. Matrixnorm [8] BANACHsche Räume.

In M_n wird eine Halbordnung $A \leq B$ erklärt, durch a_{jk}, b_{jk} reell, $0 \leq a_{jk} < b_{jk}$ $(j,k = 1,\ldots,n)$.

Mit $|A|$ bezeichnen wir die aus den Beträgen der Elemente von A gebildete Matrix.

Eine Matrixnorm heißt:

absolut, wenn $\| |A| \| = \|A\|$ . . . . . (6,1) ,

monoton, wenn aus $A \leq B$ folgt $\|A\| \leq \|B\|$ . . . . . (6,2) ,

zu einer Vektornorm passend [8], wenn

$\|Ax\| \leq \|A\| \cdot \|x\|$ für alle $A \in M_n$ und alle $x \in R_n$ . . (6,3).

Das Maximum der Beträge der charakteristischen Zahlen von A wird Spektralradius $\varrho(A)$ genannt.

Die Spektralnorm

$\|A\| = + \sqrt{\varrho(\bar{A}^t A)}$, passend zu $\|x\| = \sqrt{\sum_i |x_i|^2}$ . . . . (6,4),

ist monoton [9], aber nicht absolut, wie Gegenbeispiele zeigen.

Andere gebräuchliche Matrixnormen, wie z.B.

$\|A\| = \max_j \sum_k |a_{jk}|$, passend zu $\|x\| = \max_i |x_i|$ . . . (6,5),

$\|A\| = \max_k \sum_j |a_{jk}|$, passend zu $\|x\| = \sum_i |x_i|$ . . . . (6,6)

sind monoton und absolut.

Das Iterationsverfahren

$$x_{m+1} = Lx_m + s \quad (m = 0, 1, 2, \ldots) \quad . . . . . (6,7)$$

zur Lösung von

$$x = Lx + s \quad . . . . . (6,8)$$

konvergiert genau dann [17] für beliebiges $x_0 \in R_n$ gegen die dann eindeutig existierende Lösung von (6,8), wenn $\varrho(L) < 1$ ist. Es existiert dann [8] die Matrix $(E-L)^{-1}$, und aus (6,7),(6,8) folgt

$$x - x_{m+1} = (E-L)^{-1} L(x_{m+1} - x_m) \quad . . . . . (6,9)$$

und daraus (mit zur Vektornorm passender Matrixnorm) die Fehlerabschätzung

$$\|x_{m+1} - x\| \leq \|(E-L)^{-1} L\| \, \|x_{m+1} - x_m\| \quad . . . . . (6,10).$$

Ein bekanntes hinreichendes Kriterium von J. WEISSINGER [25]

für $\varrho(L) < 1$ ist

$$\| L \| < 1 \quad \ldots\ldots\ldots\ldots\ldots\ldots (6,11).$$

Unter dieser Voraussetzung $\| L \| < 1$ kann die Fehlerschranke in (6, 10) weiter abgeschätzt werden:

$$\| x_{m+1} - x \| \leq \frac{\| L \|}{1 - \| L \|} \| x_{m+1} - x_m \| \quad \ldots\ldots\ldots (6,12).$$

2. Ein neues hinreichendes Konvergenzkriterium und eine neue Fehlerabschätzung.

Zur Lösung eines linearen Gleichungssystems

$$x = Ax + r \quad \ldots\ldots\ldots\ldots\ldots (6,13),$$

dessen Koeffizientenmatrix $A = A_L + A_R$ Summe einer linken unteren Dreiecksmatrix A_L (mit Elementen a_{jk}, $j>k$, und sonst 0) und einer rechten oberen Dreiecksmatrix A_R (mit Elementen a_{jk}, $j \leq k$, und sonst 0) ist, berechnet man bei der Iteration in Einzelschritten, ausgehend von einem beliebigen Vektor x_0, eine Folge von Vektoren nach der Vorschrift

$$x_{m+1} = A_L x_{m+1} + A_R x_m + r \quad \ldots\ldots\ldots\ldots (6,14).$$
$$(m = 0, 1, 2, \ldots)$$

Da die n-te Potenz der Dreiecksmatrix A_L gleich der Nullmatrix und deshalb

$$(E - A_L)(E + A_L + A_L^2 + \ldots + A_L^{n-1}) = E$$

ist, existiert (als Summe einer endlichen geometrischen Reihe) die zu $E - A_L$ inverse Matrix

$$(E - A_L)^{-1} = E + A_L + A_L^2 + \ldots + A_L^{n-1} \quad \ldots\ldots\ldots (6,15).$$

Das Iterationsverfahren in Einzelschritten (6, 14) zur Lösung von (6, 13) kann daher als ein Iterationsverfahren der Art (6, 7) zur Lösung von (6, 8) mit

$$L = (E-A_L)^{-1}A_R \quad \ldots\ldots\ldots\ldots \quad (6,16),$$

$$s = (E-A_L)^{-1}r \quad \ldots\ldots\ldots\ldots \quad (6,17)$$

aufgefaßt werden. Aus (6,16), (6,15) ergibt sich

$$\begin{aligned}\|L\| &= \|(E+A_L+A_L^2+\ldots+A_L^{n-1})A_R\| \\ &\leq (1+\|A_L\|+\|A_L\|^2+\ldots+\|A_L\|^{n-1})\|A_R\| \\ &= C = \begin{cases} \dfrac{1-\|A_L\|^n}{1-\|A_L\|}\cdot\|A_R\| & \text{falls } \|A_L\| \neq 1 \\ n\cdot\|A_R\| & \text{falls } \|A_L\| = 1 \end{cases}\end{aligned}$$

und aus $\|L\| \leq C$ nach (6,11) das

Konvergenzkriterium: (6,18)

$C < 1$ ist hinreichend dafür, daß die Iteration in Einzelschritten (6,14) gegen die dann eindeutig existierende Lösung von (6,13) konvergiert,

und daraus nach (6,12) die

Fehlerabschätzung:

$$\|x_{m+1}-x\| \leq \frac{C}{1-C}\|x_{m+1}-x_m\| \quad \ldots\ldots\ldots\ldots \quad (6,19).$$

$$(m = 0,1,2,\ldots)$$

3. Verallgemeinerungen der Fehlerabschätzungen von L. COLLATZ und W. DÜCK.

Die Darstellung (6,7), (6,16) ermöglicht die Herleitung schon bekannter Fehlerabschätzungen von W. DÜCK [7] bzw. L. COLLATZ [5], [6], und zwar allgemeiner als bisher für beliebige bzw. für monotone, absolute Matrixnormen.

Aus $\|A\| < 1$ folgt die Existenz [8] der Matrix $(E-A)^{-1}$ und damit die Existenz und Eindeutigkeit der Lösung x von (6,13).

Wegen

$$(E-A)^{-1}A_R = \left\{E-A_L-A_R\right\}^{-1}A_R$$
$$= \left\{\left[E-A_L\right]\left[E-(E-A_L)^{-1}A_R\right]\right\}^{-1}A_R$$
$$= \left[E-(E-A_L)^{-1}A_R\right]^{-1}\left[E-A_L\right]^{-1}A_R$$
$$= (E-L)^{-1}L$$

existiert dann auch die Matrix $(E-L)^{-1}$, es gilt

$$\| (E-L)^{-1}L\| = \| (E-A)^{-1}A_R\| \leq \frac{\| A_R\|}{1-\| A\|}$$

und nach (6, 10) die

Fehlerabschätzung von W. DÜCK:

$$\| x_{m+1}-x\| \leq \frac{\| A_R\|}{1-\| A\|}\, \| x_{m+1}-x_m\| \quad \ldots\ldots\ldots \quad (6,20).$$
$$(m = 0, 1, 2, \ldots)$$

Diese Fehlerabschätzung läßt sich, falls die gewählte Matrixnorm monoton und absolut ist, d.h. falls aus

$$|A_R| \leq |A_L| + |A_R| = |A_L + A_R| = |A|$$

die Gültigkeit von

$$\| A_R\| = \| |A_R| \| \leq \| |A| \| = \| A\|$$

folgt, (vergröbern und) vereinfachen zur

Fehlerabschätzung von L. COLLATZ:

$$\| x_{m+1}-x\| \leq \frac{\| A\|}{1-\| A\|}\, \| x_{m+1}-x_m\| \quad \ldots\ldots\ldots \quad (6,21).$$
$$(m = 0, 1, 2, \ldots)$$

Unter der Voraussetzung $\| A\| < 1$ wurde die Konvergenz des Iterationsverfahrens bisher nicht untersucht; es gibt aber ein weiteres hinreichendes

Konvergenzkriterium von J. WEISSINGER [25]:

. . . . . (6,22)

Ist die gewählte Matrixnorm monoton und absolut und $\| A\| < 1$, so konvergiert die Iteration in Einzelschritten (6, 14) gegen die dann eindeutig existierende Lösung von (6, 13).

Für die nicht absolute Spektralnorm nennen wir ein Beispiel mit $\| A \|$ Spektral < 1, bei dem die Iteration in Einzelschritten (6, 14) divergiert:

$$A = \begin{pmatrix} 0 & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & 0 \end{pmatrix}$$

$$\| A \| \text{ Spektral} = + \sqrt{\frac{3}{4}}, \quad \varrho(L) = 1.$$

4. Ein neues hinreichendes Konvergenzkriterium und eine neue Fehlerabschätzung bei Matrizen A 'von strenger property (A)'.

Setzt man die Matrix A als 'von strenger property (A)' (2,6) voraus, d.h.

$$A = \left(\begin{array}{c|c} 0 & A_R \\ \hline A_L & 0 \end{array} \right) \begin{matrix} \uparrow n_1 \\ \downarrow \\ \uparrow n_2 \\ \downarrow \end{matrix}$$

$\leftarrow n_1 \rightarrow \quad \leftarrow n_2 \rightarrow$

so folgt aus der speziellen Rechtecksform von A_L

$$A_L^2 = 0 \quad \ldots\ldots\ldots\ldots \quad (6,23)$$

sowie $$(E - A_L)^{-1} = E + AL \quad \ldots\ldots\ldots\ldots \quad (6,24)$$

und das Iterationsverfahren in Einzelschritten (6, 14) zur Lösung von (6, 13) kann daher als ein Iterationsverfahren der Art (6, 7) zur Lösung von (6, 8) mit

$$L = (E + A_L)A_R = \left(\begin{array}{c|c} 0 & A_R \\ \hline 0 & A_L A_R \end{array} \right) \quad \ldots\ldots \quad (6,25),$$

$$s = (E + A_L)r \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (6,26)$$

aufgefaßt werden.

Wir zerlegen nun alle Vektoren in den Vektor$^{(1)}$ ihrer ersten

n_1 und den Vektor$^{(2)}$ der übrigen n_2 Komponenten und können dann das Einzelschrittverfahren schreiben als ein Iterationsverfahren :

$$\left.\begin{aligned} x^{(1)}_{m+1} &= L_1 x^{(2)}_m + s^{(1)} \\ x^{(2)}_{m+1} &= L_2 x^{(2)}_m + s^{(2)} \end{aligned}\right\} \quad \ldots\ldots \quad (6,27)$$

zur Lösung von

$$\left.\begin{aligned} x^{(1)} &= L_1 x^{(2)} + s^{(1)} \\ x^{(2)} &= L_2 x^{(2)} + s^{(2)} \end{aligned}\right\} \quad \ldots\ldots \quad (6,28)$$

mit

$$\begin{aligned} L_1 &= A_R \quad (n_1 \times n_2 \text{ Matrix}) \\ L_2 &= A_L A_R \quad (n_2 \times n_2 \text{ Matrix}) \end{aligned} \quad \ldots\ldots \quad (6,29).$$

Aus (6,27), (6,29) ergibt sich nach (6,11) für Matrizen A 'von strenger property (A)' das

Konvergenzkriterium: $\ldots\ldots\ldots$ (6,30)

$\| A_L A_R \| < 1$ ist hinreichend dafür, daß die Iteration in Einzelschritten (6,14) gegen die dann eindeutig existierende Lösung von (6,13) konvergiert,

und daraus nach (6,12) und auf Grund von

$$x^{(1)}_{m+1} - x^{(1)} = A_R (x^{(2)}_m - x^{(2)})$$

die

Fehlerabschätzung: $\ldots\ldots\ldots\ldots$ (6,31)

$$\| x^{(1)}_{m+1} - x^{(1)} \| \leq \frac{\| A_R \| \, \| A_L A_R \|}{1 - \| A_L A_R \|} \| x^{(2)}_m - x^{(2)}_{m-1} \|$$

$$(m = 1, 2, 3, \ldots)$$

$$\| x^{(2)}_{m+1} - x^{(2)} \| \leq \frac{\| A_L A_R \|}{1 - \| A_L A_R \|} \| x^{(2)}_{m+1} - x^{(2)}_m \|$$

$$(m = 0, 1, 2, \ldots)$$

(Die Theorie der Matrix und Vektornormen ist auch auf die Multiplikation einer $n_1 \times n_2$ Rechtecksmatrix mit einem Vektor

von n_2 Komponenten anwendbar.)

Eine Vereinfachung, aber auch Vergrößerung von (6,30), (6,31) ist durch Ersetzung von $\| A_L A_R \|$ durch $\| A_L \| \cdot \| A_R \|$ möglich.

5. Vergleich der Konvergenzkriterien und Fehlerabschätzungen.

a) Allgemeine Matrizen A.

Die Fehlerabschätzung (6,19) und das zugehörige Konvergenzkriterium (6,18) sind beide für beliebige Matrixnormen gültig, das zur Fehlerabschätzung von W. DÜCK (6,20) oder L. COLLATZ (6,21) gehörige Konvergenzkriterium von J. WEISSINGER (6,22) jedoch nur für monotone, absolute Matrixnormen.

An Stelle der Voraussetzung $\| A \| < 1$ von (6,20), (6,21), (6,22) tritt bei (6,18), (6,19) die Voraussetzung $C < 1$, die auch für $\| A \| > 1$ erfüllt sein kann.

Die Bedingung $C < 1$ läßt zu, daß in der Matrix $E - A_L - A_R$ Elemente von A_L betragsmäßig überwiegen, falls die Elemente von A_R hinreichend kleinen Betrages sind; insbesondere zeigt die Fehlerabschätzung (6,19) für beliebige A_L und $A_R = 0$ die exakte Lösung nach einem Iterationsschritt an.

Für $\| A \| < 1$ kann, wie Beispiele zeigen, die Fehlerabschätzung (6,19) besser oder schlechter sein als die Fehlerabschätzungen von L. COLLATZ und W. DÜCK.

Die für die Matrixnorm der maximalen Zeilenbetragssumme (mit Gewichten) gültige Fehlerabschätzung von H. SASSENFELD [19] (und deren Verallgemeinerung von J. SCHRÖDER [20]) ist bezüglich dieser speziellen Norm besser oder genau so gut wie die Fehlerabschätzungen (6,19), (6,20), (6,21).

b) Matrizen A 'von strenger property (A)'.

Da die unter a) erwähnten Konvergenzkriterien und Fehlerabschätzungen auch unter Voraussetzung der 'strengen property (A)' anwendbar sind, gilt das im Fall a) Gesagte auch im Fall b). Hinzu kommen nun das Konvergenzkriterium (6,30) und die Fehlerabschätzung (6,31), die wieder beide für beliebige Matrixnormen gültig sind. An Stelle der Voraussetzungen $\| A\| < 1$ von (6,20), (6,21), (6,22) und $C < 1$ von (6,18), (6,19) tritt die schwächere Voraussetzung $\| A_L A_R\| < 1$. Für beliebige Matrixnormen gilt $\| A_L A_R\| \leq C$. Noch besser als die Bedingung $C < 1$ zeigt $\| A_L A_R\| < 1$, daß das Einzelschrittverfahren konvergiert, wenn eine der beiden Matrizen A_L, A_R hinreichend kleine Komponentenbeträge besitzt und daß für beliebige A_L und $A_R = 0$ oder beliebige A_R und $A_L = 0$ die exakte Lösung in einem Iterationsschritt erreicht wird.

Die Fehlerabschätzung (6,31) ist besser oder genau so gut wie die Fehlerabschätzung (6,19) und die Fehlerabschätzungen von L. COLLATZ (6,21) und W. DÜCK (6,20) und für die spezielle Matrixnorm der maximalen Zeilenbetragssumme besser oder genau so gut wie die Fehlerabschätzung von H. SASSENFELD [19].

§ 7. BEISPIELE .

Beispiele zu § 2.

1.) Matrizen A 'von strenger property (A)' treten auf [10] beim gewöhnlichen Differenzenverfahren zur Lösung von 1. Randwertaufgaben elliptischer partieller Differentialgleichungen (DIRICHLETsches Problem), in denen die mittleren Terme fehlen (in 2 Dimensionen u_{xy}).

Symmetrische Matrizen A 'von strenger property (A)' treten auf bei dem speziellen Typ (c konstant):

$$\left.\begin{array}{l} \Delta u + c \cdot u + f(x_1, \ldots, x_N) = 0 \\ \text{im N-dimensionalen Bereich B ,} \\ u = g(x_1, \ldots, x_N) \\ \text{auf dem Rand } \Gamma \text{ von B.} \end{array}\right\} \quad \ldots\ldots \quad (7,1).$$

Das gewöhnliche Differenzenverfahren (nicht Mehrstellenverfahren) liefert dann für jeden der n inneren Gitterpunkte des mit einem Quadratnetz (N = 2 als Beispiel) überdeckten Bereiches B eine lineare Gleichung zur Bestimmung des Wertes $u_{i,j}$ aus den vier Nachbarwerten:

$$u_{i,j} - \alpha u_{i+1,j} - \alpha u_{i,j+1} - \alpha u_{i-1,j} - \alpha u_{i,j-1} = \beta_{i,j} \quad \ldots\ldots\ldots \quad (7,2),$$

$$\alpha = \frac{1}{4 - h^2 c} \quad , \quad \beta_{i,j} = \frac{h^2}{4 - h^2 c} f_{i,j} \quad ,$$

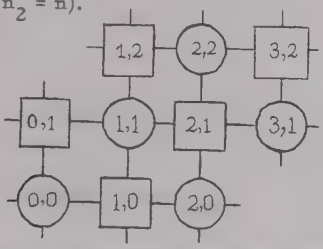
Maschenweite $h \neq \frac{2}{\sqrt{c}}$.

Stellt man diese Gleichung erst für alle n_1 Unbekannten $u_{i,j}$ mit gerader Indexsumme i+j und anschließend für alle n_2 Unbekannten mit ungerader Indexsumme auf [26] , so erhält man ein lineares Gleichungssystem

$$(E - A)\,\tilde{u} = r \quad \ldots\ldots\ldots\ldots \quad (7,3),$$

dessen n × n Teilmatrix $A = A_L + A_R$ nur aus Elementen 0 oder α besteht, symmetrisch ist und 'von strenger property (A)' ($n_1 + n_2 = n$).

Abb. 3



Zum Nachweis der Symmetrie und der 'strengen property (A)' denken wir uns die n Unbekannten $u_{i,j}$ mit nur einem Index k als u_k numeriert, der in ◯ die n_1 Zahlen 1 bis n_1 und in □ die n_2 Zahlen $n_1 + 1$ bis n durchläuft, und weisen darauf hin, daß in der k-ten Zeile und l-ten Spalte von A ein Element α genau dann auftritt, wenn u_k und u_l Nachbarwerte sind.

2.) In einem Spezialfall von (7, 1), bei der Randwertaufgabe

$$\left.\begin{array}{l} \Delta u + f(x,y) = 0 \\ \text{in einem Rechtecksbereich } B, \\ u = g(x,y) \\ \text{auf dem Rand } \Gamma \text{ von } B \end{array}\right\} \quad \ldots\ldots\ldots \quad (7,4)$$

(analog in höheren Dimensionen [26])

und unter der Voraussetzung, daß die Rechtecksseite a das (n_a+1)-fache und die Rechtecksseite b das (n_b+1)-fache der Maschenweite h beträgt, sind die n Eigenwerte der Matrix A bekannt [26]:

$$\lambda_{p,q} = \frac{1}{2}\left(\cos\frac{p}{n_a+1}\pi + \cos\frac{q}{n_b+1}\pi\right) \quad \ldots\ldots\ldots \quad (7,5)$$

$(p=1,\ldots,n_a, \quad q=1,\ldots,n_b, \quad n_a \cdot n_b = n)$

In diesem Spezialfall ist es möglich, nach der Theorie von D. YOUNG [26] § 2 (2,12) den optimalen Overrelaxationsfaktor ω_b direkt zu berechnen:

$$\omega_b = \frac{2}{1+\sqrt{1-\lambda_{1,1}^2}} \qquad \ldots\ldots\ldots (7,6)$$

Beispiele zu § 3

1. Ein lineares Gleichungssystem (n=20) mit symmetrischer Matrix A 'von strenger property (A)'

Die Lösung der Randwertaufgabe (Temperaturverteilung in einem Rechtecksbereich)

$$\Delta u = 0 \quad \text{in} \quad |x| < 3\,, \quad |y| < 2{,}5$$

$$u = -6y+15 \text{ für } x = +3\,, \quad u = 0 \text{ für } x = -3$$

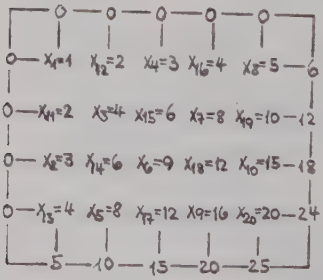
$$u = 5x + 15 \text{ für } y = -2{,}5, u = 0 \text{ für } y = +2{,}5$$

nach den Differenzenverfahren mit einem Quadratgitter von $4 \times 5 = 20$ inneren Punkten führt auf die Lösung eines linearen Gleichungssystems $x = Ax + r$ (n=20) mit symmetrischer Matrix A 'von strenger property (A)'.

$$A = \begin{pmatrix} 0 & | & A_R \\ - & + & - \\ A_L & | & 0 \end{pmatrix}, \quad A_R = A_L' = \begin{pmatrix} \alpha & \cdot & \alpha & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \alpha & \cdot & \alpha & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \alpha & \cdot & \alpha & \cdot & \alpha & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \alpha & \cdot & \alpha & \cdot & \alpha & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \alpha & \cdot & \alpha & \cdot & \alpha & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \alpha & \cdot & \alpha & \cdot & \alpha & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \alpha & \cdot & \alpha & \cdot & \alpha & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \alpha & \cdot & \alpha & \cdot & \alpha \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \alpha & \cdot & \alpha & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \alpha & \cdot & \alpha \end{pmatrix}, \quad r^{(1)} = \begin{pmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ 2{,}5 \\ \cdot \\ \cdot \\ 1{,}5 \\ 5 \\ 4 \end{pmatrix}, \quad r^{(2)} = \begin{pmatrix} \cdot \\ \cdot \\ 1{,}25 \\ \cdot \\ \cdot \\ \cdot \\ 3{,}75 \\ \cdot \\ 3 \\ 12{,}75 \end{pmatrix}.$$

($\cdot = 0$, $\alpha = 0{,}25$)

Bei diesem speziellen Beispiel sind die exakten Lösungen des Gleichungssystems ganze Zahlen und die Eigenwerte von A nach Formel (7,5) direkt zu berechnen:



Eigenwerte von A:

$\pm 0{,}8375\ 2120$	$\pm 0{,}2785\ 0372$
$\pm 0{,}6545\ 0850$	$\pm 0{,}1545\ 0851$
$\pm 0{,}5875\ 2109$	$\pm 0{,}1545\ 0851$
$\pm 0{,}4045\ 0831$	$\pm 0{,}0954\ 9109$
$\pm 0{,}4045\ 0831$	$\pm 0{,}0285\ 0457$

2. Iterative Bestimmung des ersten Eigenwertes $\eta_1(1)$ der Matrix L(1) nach dem δ^2-Verfahren (3, 16).

$L(1)=A_R+A_L A_R$ ist die Matrix des Einzelschrittverfahrens $x_{m+1} = L(1)x_m+r$ (m=0, 1, 2, ...) zur Lösung des Gleichungssystems (n=20) x=Ax+**r**, $A=A_L+A_R$ von Seite 50 . ($x_0 = 0$)

m	1. Komponente von x_m	$Q\delta x_m$	$PQ\delta x_m$
1	0	0,509213600	0,700751380
2	0	0,635500757	0,701668504
3	0,064453125	0,678527277	0,701388622
4	0,214843750	0,693581620	0,701403689
5	0,392181305	0,698722255	0,701424733
6	0,549895723	0,700485240	0,701430566
7	0,674547030	0,701098940	0,701457188
8	0,767731680	0,701317690	0,701405150
9	0,825441380	0,701397130	0,701511340
10	0,858389596	0,701425330	0,702047010
11	0,918279060	0,701437650	0,701754390
12	0,942560350	0,701436780	0,701530610
13	0,959660330	0,701436480	0,703846150
14	0,971686343	0,701440100	0,700920250
15	0,980128910	0,701446320	0,701579380
16	0,986057580	0,701446020	0,701634e60
17	0,990218900	0,701433690	0,701363340
18	0,993138460	0,701450720	0,701457730
19	0,995156750	0,701421540	0,701464080
20	0,996623710	0,701478110	0,701390110
21	0,997631680	0,701403540	0,699387755
22	0,998338675	0,701442630	0,701386310
23	0,998833475	0,701481230	0,701533670
24	0,999181268	0,701668141	0,701331170
25	0,999426873	0,701213910	0,701330420
26	0,999595788	0,701374430	0,701409350
27	0,999717196	0,701342530	0,701468750
28	0,999806221	0,701466730	0,701474470
29	0,999856133	0,701503100	0,701374170
30	0,999902273	0,701175950	0,705718820
31	0,999931683	0,701655830	0,701239b30
32	0,999952223	0,702085350	0,700037320
33	0,999966055	0,702964930	0,703286777
34	0,999973056	0,704219690	0,705370870
35	0,999982355	0,700027488	0,693277100
36	0,999986850	0,682812160	0,698264300
37	0,999991584	0,708927340	0,699325932
38	0,999994400	0,699090598	0,699164000
39	0,999996311	0,699361265	0,695290999
40	0,999997730	0,686396990	0,699841150
41	0,999998213	0,753288820	0,695327710
42	0,999999871	0,487335764	0,601190530
43	0,999999915	0,765185590	0,427969760
44	0,999999943	0,364928910	0,129128980
45	0,999999963	0,439754100	—
46	0,999999973	0,590601570	—
47	0,999999980	—	—
48	0,999999986	—	—
	(1,000000000)	(0,701441760)	(0,701441760)

Einfluß von Rundungsfehlern

Da die Matrix A symmetrisch ist und 'von strenger property (A)', kann nach Satz (4, 2) und den Bemerkungen von Seite 38 das δ^2-Verfahren

$$PQ\delta x_m = \eta_1(1) + o\left|\frac{\eta_2(1)^2}{\eta_1(1)^2}\right|^m + o\left|\frac{\eta_3(1)}{\eta_1(1)}\right|^m \quad \text{vgl. (3, 16)}$$

$$(|\eta_1(1)| > |\eta_2(1)| > |\eta_3(1)| \geq \ldots \geq |\eta_n(1)|)$$

angewandt werden, ohne daß die in Klammern gesetzten Voraussetzungen vorher überprüft zu werden brauchen.

Die Ordnung des Restgliedes läßt sich (bei dieser speziellen Matrix) aus den Eigenwerten von A (Seite 50) berechnen:

$$\left|\frac{\eta_2(1)^2}{\eta_1(1)^2}\right| = 0,37 \quad < \quad \left|\frac{\eta_3(1)}{\eta_1(1)}\right| = \underline{0,49}$$

Es wird ausdrücklich darauf hingewiesen, daß der Operator $PQ\delta$ auf Vektoren x_m angewandt wird, aus Platzmangel aber nicht alle Komponenten von x_m, sondern nur die 1. Komponente von x_m - in dieser und allen folgenden Tabellen - angegeben werden können.

3. Iterative Bestimmung des zweiten Eigenwertes $\eta_2(1)$ der Matrix L(1) nach dem Verfahren (3, 19) unter Verwendung des in 2) berechneten ersten Eigenwertes $\eta_1(1)$.

$$PQ\,\delta\,\delta\,\eta_1(1)^x{}_m = \eta_2(1) + o\left|\frac{\eta_3(1)^2}{\eta_2(1)^2}\right|^m + o\left|\frac{\eta_4(1)}{\eta_2(1)}\right|^m \quad \text{vgl. (3, 19)}$$

$$(\eta_1(1) > |\eta_2(1)| > |\eta_3(1)| > |\eta_4(1)| \geq \ldots \geq |\eta_n(1)|)$$

Zur Anwendung dieses Verfahrens gelten die gleichen Bemerkungen wie oben beim δ^2-Verfahren.

Die Ordnung des Restgliedes läßt sich (bei dieser speziellen Matrix) aus den Eigenwerten von A (Seite 50) berechnen:

$$\left|\frac{\eta_3(1)^2}{\eta_2(1)^2}\right| = \underline{0,65} \quad > \quad \left|\frac{\eta_4(1)}{\eta_2(1)}\right| = 0,38$$

m	1.Komponente von x_m	1. Komponente von $\delta\eta_1(1)^{x_m}$	$Q\delta\delta\eta_1(1)^{x_m}$	$PQ\delta\delta\eta_1(1)^{x_m}$
1	0	0	0,25460855	0,40218718
2	0	0,06445312	0,36012865	0,41067107
3	0,00445312	0,16963364	0,39020092	0,42239875
4	0,21484375	0,24148867	0,40238073	0,43206202
5	0,39218905	0,27479945	0,40995314	0,42816552
6	0,54989723	0,28882615	0,41555942	0,43179413
7	0,67454703	0,29457622	0,41948576	0,42675932
8	0,76773168	0,29692232	0,42244477	0,40490798
9	0,83544138	0,29788249	0,42415976	0,42572804
10	0,88389596	0,29827752	0,42613079	0,42606202
11	0,91827906	0,29844107	0,42562308	0,43085228
12	0,94256035	0,29850914	0,43067594	0,43067736
13	0,95966033	0,29853760	0,43084795	0,45785062
14	0,97166343	0,29854957	0,40589022	0,39440131
15	0,98012891	0,29855453	0,35786477	0,38997729
16	0,98605788	0,29855673	0,51059281	-2,19317389
17	0,99021890	0,29855757	0,06305871	—
18	0,99313846	0,29855796	0,51500000	—
19	0,99516075	0,29855816	—	—
20	0,99662371	0,29855819	—	—
21	0,99763168	—	—	—

(1,00000000)(0,29855824) (0,42838138)(0,42838138)

(Einfluß von Rundungsfehlern.)

Bei Rechnung mit mehr als 8 Stellen (für x_m) kann $\eta_2(1)$ nach diesem Verfahren auch auf mehr als 2 gültige Stellen berechnet werden.

Beispiele zu § 5

1. Lösung des Gleichungssystems (n=20) von Seite 50 nach der Iterativen Methode (5,2).

a) Einzelschrittverfahren (x_0=0), δ^2-Verfahren und Bestimmung von ω_b auf 3 Stellen

m	1. Komponente von x_m	$Q\delta x_m$	$PQ\delta x_m$	ω_b
1	0	0,50921360	0,70075138	1,29279 46
2	0	0,65550757	0,70168584	1,29335092
3	0,06445312	0,67852727	—	—
4	0,21484375	0,69358162	—	—
5	0,39218905	—	—	—
6	0,54989723	—	—	—

(1,00000000)(0,70144176)(0,70144176) (1,2933224)

b) Konvergenzbeschleunigung (3,25)
mit $\eta_1(1) \cong$ 0,70168584 (0,70144176) aus a)

1. Komponente von $\tilde{x}$	2. Komponente von $\tilde{x}$	3. Komponente von $\tilde{x}$
0,92413693	2,92926600	3,88954000

(1,00000000) (3,00000000) (4,00000000)

c) Overrelaxation mit $x_0 = \tilde{x}$ aus b) und
$\omega_b \cong$ 1,294 (1,2933224) aus a)

m	1. Komponente von x_m	2. Komponente von x_m	3. Komponente von x_m	Iterat. insges.
1	0,98267171	2,98064500	3,97561885	7
2	1,00054403	3,00413544	4,00839344	8
3	1,00021915	3,00278655	4,00352144	9
4	0,99998297 0	3,00016606	3,99992400	10
5	0,99998511 8	2,99981744	3,99980622	11
6	1,00000118	2,99999795	4,00001853	12
7	1,00000154	3,00000210	4,00000279	13
8	1,00000023	3,00000070	4,00000059	14
9	0,99999998 0	2,99999995	4,00000002	15
10	1,00000001	2,99999998	4,00000002	16
11	1,00000001	3,00000003	4,00000006	17

(1,00000000) (3,00000000) (4,00000000)

7 gültige Stellen

2. Vergleich: Lösung des Gleichungssystems (n=20) von Seite 50 nach der Iterativen Methode (5,2), nach dem Einzelschrittverfahren und - mit Overrelaxation bei als bekannt vorausgesetztem Faktor $\omega_b = 1{,}2933224$.

($x_0 = 0$)

m	1. Komponente von x_m Methode (5,2)	1. Komponente von x_m Einzelschrittv.	1. Komponente von x_m Overrel. mit ω_b
1	0	0	0
2	0	0	0
3	0,064453125	0,064453125	0,217197711
4	0,214843750	0,214843750	0,732784422
5	0,392189005	0,392189005	0,958317234
6	0,549897230	0,549897230	0,967058630
	0,924136993		
7	0,982671710	0,674547030	0,991832480
8	1,005440300	0,767731680	0,997073800
9	1,002191500	0,835441380	0,999156599
10	0,999982970	0,883895960	0,999706960
11	0,999985118	0,918279060	0,999910450
12	1,000001180	0,942560350	0,999969310
13	1,000001540	0,958666330	0,999989640
14	1,000000023	0,971683430	0,999995752
15	0,999999960	0,980126910	0,999999319
16	1,000000001	0,985057880	0,999999985
17	1,000000001	0,990218900	1,000000001
18		0,993138460	1,000000001
19		0,995186750	
20		0,996623711	
21		0,997631680	
22		0,998338750	
23		0,998834750	
24		0,999182680	
25		0,999426730	
26		0,999595788	
27		0,999717960	
28		0,999802210	
29		0,999866133	
30		0,999902730	
31		0,999931830	
32		0,999952230	
33		0,999966550	
34		0,999976560	
35		0,999983550	
36		0,999988500	
37		0,999991980	
38		0,999994400	
39		0,999996110	
40		0,999997300	
41		0,999998130	
42		0,999998710	
43		0,999999150	
44		0,999999430	
45		0,999999630	
46		0,999999730	
47		0,999999860	
48		0,999999960	

7 gültige Stellen

3. Ein lineares Gleichungssystem (n=50) mit symmetrischer Matrix A 'von strenger property (A)'.

Die Lösung der Randwertaufgabe (Temperaturverteilung in einem Rechtecksbereich)

$$\Delta u = 0 \quad \text{in} \quad |x| < 5{,}5 \;,\; |y| < 3 \;,$$

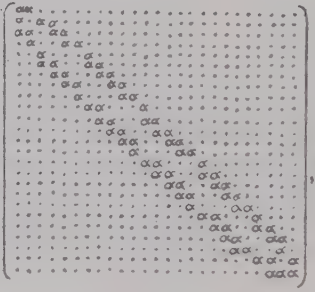
$$u = -11x+33 \text{ für } x = +5{,}5 \;,\; u = 0 \text{ für } x = -5{,}5 \;,$$

$$u = 6x+33 \text{ für } y = -3 \;,\; u = 0 \text{ für } y = +3$$

nach dem Differenzenverfahren mit einem Quadratgitter von $5 \times 10 = 50$ inneren Punkten führt auf die Lösung eines linearen Gleichungssystems $x = Ax + r$ (n=50) mit symmetrischer Matrix A 'von strenger property (A)'.

$$A = \begin{pmatrix} 0 & A_R \\ A_L & 0 \end{pmatrix}, \quad A_R = \begin{pmatrix} \alpha & \alpha & & & \\ \alpha & \alpha & \alpha & & \\ & \alpha & \alpha & \alpha & \\ & & \alpha & \alpha & \alpha \\ & & & \alpha & \alpha \end{pmatrix}$$

$(A_R = A_L^T, \alpha = 0{,}25)$



$$r^{(1)} = \begin{pmatrix} \\ \\ 4{,}5 \\ \\ 4{,}5 \\ \\ 7{,}5 \\ \\ 10{,}5 \\ \\ 13{,}5 \\ 11 \end{pmatrix}, \quad r^{(2)} = \begin{pmatrix} \\ \\ 2 \\ \\ 6 \\ \\ 9 \\ \\ 2{,}75 \\ 2 \\ 4{,}25 \\ 28{,}75 \end{pmatrix}$$

Bei diesem speziellen Beispiel sind die exakten Lösungen des Gleichungssystems ganze Zahlen und die Eigenwerte von A nach Formel (7,5) direkt zu berechnen.

Betragsmäßig größte Eigenwerte von A: $\pm\, 0{,}9127\,5920$

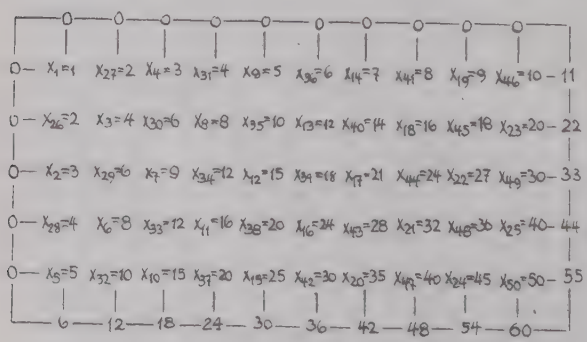


Abb. 4

4. Lösung des Gleichungssystems (n=50) von Seite 56 nach der Iterativen Methode (5,2).

a) Einzelschrittverfahren (x_0=0), δ^2-Verfahren und Bestimmung von ω_b auf 3 Stellen.

m	1. Komponente von x_m	$Q\delta x_m$	$PQ\delta x_m$	ω_b
1	0	0,59959844	0,80452045	1,36868368
2	0	0,69838869	0,82316264	1,40793447
3	0,00585937	0,75853051	0,82538057	1,41105617
4	0,04394531	0,78966836	—	—
5	0,11089326	0,80631901	—	—
6	0,19260405	—	—	—
7	0,27842326	—	—	—

(1,00000000) (0,83312936) (0,83312936) (1,#199523)

b) Konvergenzbeschleunigung (3, 25)

mit $\eta_1(1) \cong 0,82538057$ (0,83312936) aus a)

1. Komponente von $\tilde{x}$	2. Komponente von $\tilde{x}$	3. Komponente von $\tilde{x}$
0,68867047	2,55840722	3,06616116

(1,00000000) (3,00000000) (4,00000000)

c) Overrelaxation mit $x_0 = \tilde{x}$ aus b) und $\omega_b \cong 1,42$ (1,41995229) aus a)

m	1. Komponente von x_m	2. Komponente von x_m	3. Komponente von x_m	Iterat. insges.
1	0,79250699	2,65825122	3,36684445	8
2	0,90662293	2,80195988	3,70815501	9
3	0,96901651	2,92043460	3,95053450	10
4	0,98525005	2,96791396	3,97[illegible]	11
5	1,00257251	3,00770220	4,00622394	12
6	1,00154577	3,00455116	4,00318501	13
7	0,99998630	2,99955566	3,99902011	14
8	0,99993928	2,99841396	3,99830557	15
9	0,99995732	2,99852954	3,99837112	16
10	0,99998094	2,99956753	3,99942465	17
11	0,99999416	2,99999096	3,99981656	18
12	0,99999836	2,99999695	3,99995600	19
13	0,99999734	2,99999686	3,99995471	20
14	0,99999778	2,99999761	3,99996526	21
15	0,99999870	2,99999886	3,99998422	22
16	0,99999959	2,99999995	3,99999869	23
17	1,00000000	3,00000001	4,00000002	24

(1,00000000) (3,00000000) (4,00000000)

7 gültig. Stellen

5. Vergleich: Lösung des Gleichungssystems (n=50) von Seite 56 nach der Iterativen Methode (5,2), nach dem Einzelschrittverfahren und - mit Overrelaxation bei als bekannt vorausgesetztem Faktor $\omega_b = 1,41995229$. ($x_0 = 0$)

m	1. Komponente von x_m Methode (5,2)	1. Komponente von x_m Einzelschrittverf.	1. Komponente von x_m Overrel. mit ω_b
1			
2			
3	0,0058593 7	0,0058593 7	0,0338236 2
4	0,0439453 1	0,0439453 1	0,3556612 2
5	0,1108932 6	0,1108932 6	0,6982754 5
6	0,1926040 8	0,1926040 8	0,7940193 3
7	0,2784232 6	0,2784232 6	0,8876873 4
	0,6867604 7		
8	0,7925069 9	0,3623687 1	0,9775931 07
9	0,9066229 3	0,4413914 0	0,9869275 5
10	0,9690616 5	0,5140314 2	0,9914299 27
11	0,9956200 5	0,5796298 61	0,9980714 0
12	1,0025725	0,6383085 6	0,9984570 7
13	1,0015457	0,6900841 8	0,9990507 7
14	0,9998836 30	0,7354343 88	0,9996648 9
15	0,9993928 5	0,7746751 3	0,9998858 26
16	0,9995792 0	0,8089687 8	0,9999248 7
17	0,9998094 2	0,8385289 66	0,9999371 49
18	0,9999416 6	0,8633957 8	0,9999677 5
19	0,9999893 6	0,8848127 6	0,9999949 2
20	0,9999997 34	0,9030241 5	0,9999958 38
21	0,9999997 78	0,9184672 3	0,9999999 02
22	0,9999998 70	0,9315319 3	0,9999999 79
23	0,9999999 59	0,9425620 6	1,0000 0001
24	1,0000000	0,9518580 6	1,0000 0002
25		0,9596607 8	
26		0,9662550 6	
27		0,9717738 6	
28		0,9764021 3	
29		0,9802802 3	
30		0,9833527 38	
31		0,9862445 1	
32		0,9885167 5	
33		0,9904161 8	
34		0,9920031 8	
35		0,9933286 8	
36		0,9944354 3	
37		0,9953593 0	
38		0,9961302 3	
39		0,9967773 55	
40		0,9973101 3	
41		0,9977757 68	
42		0,9981309 1	
43		0,9984421 0	
44		0,9987015 8	
45		0,9989179 3	
⋮		⋮	
87		1,0000001	7 gültige Stellen

Beispiele zu § 6

1. Fehlerabschätzungen zum Einzelschrittverfahren bei Matrizen A ohne 'strenge property (A)'.

Es wird speziell die Vektornorm des maximalen Komponentenbetrages und die dazu passende (6,3) Matrixnorm der maximalen Zeilenbetragssumme gewählt und für das Einzelschrittverfahren (6,14) bei den beiden Beispielen (1) und (2)

(1)

$$A=\begin{pmatrix} 0 & 0 & 0{,}001 & 0 & 0 \\ 0 & 0 & 0{,}001 & 0 & 0 \\ 2 & 2 & 0 & 0{,}001 & 0{,}001 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \end{pmatrix}$$

(2)

$$A=\begin{pmatrix} 0 & 0 & 0 & 0 & 0{,}005 \\ 0 & 0 & 0 & 0 & 0{,}005 \\ 0 & 0 & 0 & 0 & 0{,}005 \\ 0 & 0 & 0 & 0 & 0{,}005 \\ 0{,}2 & 0{,}2 & 0{,}2 & 0{,}2 & 0 \end{pmatrix}$$

die in

$$\| x_{m+1}-x\| \leq q \| x_{m+1}-x_m\| \quad (m = 0,1,2,\ldots)$$

auftretende reelle nichtnegative Konstante q abgeschätzt:

(Ein Strich - bedeutet, daß das zugehörige Konvergenzkriterium versagt.)

Fehlerabschätzung	Beispiel (1) q	Beispiel (2) q
H. SASSENFELD [19]	0,013	0,053
(6,19)	2,074	0,203
W. DÜCK [7]	-	0,25
L. COLLATZ [5]	-	4

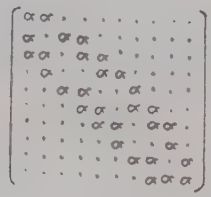
Die Fehlerabschätzung von H. SASSENFELD [19] zum Einzelschrittverfahren ist zwar schärfer als die Fehlerabschätzung (6,19), aber zu ihrer Durchführung benötigt man den Rechenaufwand einer Matrix × Vektor - Multiplikation im Gegensatz zu der numerisch sehr viel einfacheren neuen Fehlerabschätzung (6,19):

$$q = \frac{C}{1-C}\ , \quad C = \left\{ \begin{array}{ll} \dfrac{1 - \|A_L\|^n}{1 - \|A_L\|} \cdot \|A_R\| & \text{falls } \|A_L\| \neq 1 \\ n \cdot \|A_R\| & \text{falls } \|A_L\| = 1 \end{array} \right\} .$$

2. Fehlerabschätzung zum Einzelschrittverfahrer bei (symmetrischer) Matrix A 'von strenger property (A)'.

Es werden die gleichen speziellen Normen wie in 1. gewählt und das Konvergenzkriterium (6,30) und die Fehlerabschätzung (6,31) auf das Beispiel (n=20) von Seite 51 angewandt.

$$A = \left(\begin{array}{c|c} 0 & A_R \\ \hline A_L & 0 \end{array} \right) \ , \qquad A_R = A_L^T =$$

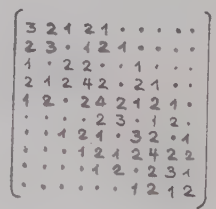


$(\cdot = 0 \ , \ \alpha = 0{,}25)$

Da die Matrix A_R zur Nebendiagonale symmetrisch ist und viele Nullen aufweist, ist $A_L \cdot A_R$ einfach zu bilden.

$(\cdot = 0)$

$$A_L A_R = \alpha^2$$



Das hinreichende Konvergenzkriterium (6,30) ist erfüllt:

$$\| A_L \cdot A_R \| = \frac{15}{16} < 1$$

und die Fehlerabschätzung (6,31):

$$\| x_{m+1}^{(1)} - x^{(1)} \| \leq \frac{\| A_R \| \, \| A_L A_R \|}{1 - \| A_L A_R \|} \| x_m^{(2)} - x_{m-1}^{(2)} \|$$

$$\| x_{m+1}^{(2)} - x^{(1)} \| \leq \frac{\| A_L A_R \|}{1 - \| A_L A_R \|} \| x_{m+1}^{(2)} - x_m^{(2)} \| \quad \ldots \text{vgl. (6,31)}$$

ergibt mit $\| A_R \| = 1$, $\| A_L A_R \| = \frac{15}{16}$:

$$\| x_{m+1}^{(1)} - x^{(1)} \| \leq 15 \| x_m^{(2)} - x_{m-1}^{(2)} \| \quad (m = 1,2,3,\cdots)$$

$$\| x_{m+1}^{(2)} - x^{(2)} \| \leq 15 \| x_{m+1}^{(2)} - x_m^{(2)} \| \quad (m = 0,1,2,\cdots) .$$

($x_m^{(1)}$ ist der Vektor der ersten $n_1 = 10$ Komponenten von x_m und $x_m^{(2)}$ ist der Vektor der übrigen $n_2 = 10$ Kompoenten von x_m.)

Das zur Fehlerabschätzung von L. COLLATZ [5] und W. DÜCK [7] gehörige Konvergenzkriterium $\| A \| < 1$ versagt bei Matrizen des Differenzenverfahrens. Das Konvergenzkriterium (6,19) und das Konvergenzkriterium von H. SASSENFELD [19] versagen, wenn (wie in diesem Beispiel) die Matrizen des Differenzenverfahrens so aufgestellt werden, daß 'strenge property (A)' vorliegt.

LITERATURVERZEICHNIS

[1] AITKEN, A. C.: The Evaluation of the Latent Roots and Latent Vektors of a Matrix, Proc. Royal Soc. Edinburgh 57 (1937) S. 268-304

[2] AITKEN, A. C.: On the Iterative Solution of a System of Linear Equations, Proc. Royal Soc. Edinburgh. Section A 63 (1950) S. 52-80

[3] ALBRECHT, J.: Fehlerabschätzungen bei Relaxationsverfahren zur numerischen Auflösung linearer Gleichungssysteme. Num. Mathem. 3 (1961) 188-201

[4] ARMS, L. J., GATES, L. D., ZONDEK, B.: A Method of Block Iteration, J. Soc. Ind. Appl. Math. 4 (1956) S. 220-229

[5] COLLATZ, L.: Fehlerabschätzung für das Iterationsverfahren zur Auflösung linearer Gleichungssysteme, ZAMM 22 (1942) S. 357-361

[6] COLLATZ, L.: Einige Anwendungen funktionalanalytischer Methoden in der praktischen Analysis, ZAMP 4 (1953) S. 327-357

[7] DÜCK, W.: Eine Fehlerabschätzung zum Einzelschrittverfahren bei linearen Gleichungssystemen, Numerische Mathematik 1 (1959) S. 73-77

[8] FADDEEVA, V. N.: Computational Methods of Linear Algebra, New York 1959, Dover Public. S. 54-64 ff.

[9] FELDMANN, H.: Ein hinreichendes Konvergenzkriterium und eine Fehlerabschätzung für die Iteration in Einzelschritten bei linearen Gleichungssystemen, ZAMM 41 (1961) S. 515-516

[10] FORSYTHE, G.E., WASOW, W.R.: Finite Difference Methods for Partial Differential Equations, New York - London (1960), 444 S.

[11] FRIEDMAN, B.: The Iterative Solution of Elliptic Difference Equations, Report NYO-7698 (1957), Institute of Mathematical Sciences, New York University, 37 S.

[12] GAUSS, C.F.: Brief an Gerling (26.12.1823) Werke Vol. 9 S. 278-281

[13] GOLUB, G.H., VARGA, R.S.: Chebyshev Semi Iterative Methods, Successive Overrelaxation Iterative Methods, and Second Order Richardson Iterative Methods, Cleveland (1960), Case Institute of Technology, Report No. 1028, 46 S.

[14] HESTENES, M.R., STIEFEL, E.: Method of Conjugate Gradients for Solving Linear Systems, J. Res. Nat. Bur, Standards, Vol. 49 (1958) S. 409-436

[15] JAKOBI, C.G.J.: Ueber eine neue Auflösungsart der bei der Methode der kleinsten Quadrate vorkommenden lineären Gleichungen, Astr. Nachr. Vol. 22 No. 523 (1845) S. 297-306

[16] KAHAN, W.: Gauss-Seidel Methods of Solving Large Systems of Linear Equations, Thesis (1958) Toronto, etwa 100 S.

[17] OLDENBURGER, R.: Infinite Powers of Matrices and Characteristic Roots, Duke, Math. J. (1940) S. 357-361

[18] RICHARDSON. L. F.: The Approximate Arithmetical Solution by Finite Differences of Physical Problems Involving Differential Equations, with an Application to the Stresses in a Masonry Dam, Philos. Trans. Roy. Soc. Lond. Ser. A, Vol. 210, S. 307-357

[19] SASSENFELD, H.: Ein hinreichendes Konvergenzkriterium und eine Fehlerabschätzung für die Iteration in Einzelschritten bei linearen Gleichungen, ZAMM 31 (1951) S. 92-94

[20] SCHRÖDER, J.: Neue Fehlerabschätzungen für verschiedene Iterationsverfahren, ZAMM 36 (1956) S. 168-181

[21] SEIDEL, L.: Ueber ein Verfahren, die Gleichungen, auf welche die Methode der kleinsten Quadrate führt, sowie lineäre Gleichungen überhaupt, durch successive Annäherung aufzulösen. Abh. Math. Phys. Kl. Bayer. Akad. Wiss., München (1874) Vol. 11(3) S. 81-108

[22] SOUTHWELL, R. V.: Relaxation Methods in Theoretical Physics, Clarendon Press, Oxford Vol. 1 (1946) Vol. 2 (1956) 522 S.

[23] STIEFEL, E. L.: Kernel Polynomials in Linear Algebra and their Numerical Applications, N. B. S. A. M. S. Vol. 49(1958) S. 1-22

[24] TEMPLE, G.: The General Theory of Relaxation Methods Applied to Linear Systems, Proc. Roy. Soc. 169 (1938) S. 476

[25] WEISSINGER, J.: Zur Theorie und Anwendung des Iterationsverfahrens, Math. Nachr. 8 (1952) S. 193-212

[26] YOUNG, D.: Iterative Methods for Solving Partial Difference Equations of Elliptic Type, Trans. Amer. Math. Soc. 76 (1954) S. 92-111

LEBENSLAUF.

Geboren wurde ich am 15. Dezember 1934 in Hamburg als Sohn des Studienassessors (heute Oberstudienrat) Dr. rer. nat. Harry Feldmann und seiner Ehefrau, der technischen Lehrerin Käthe Feldmann geb. Hack. Von 1941 bis 1943 besuchte ich die Volksschule in Hamburg, nach unserer Ausbombung 1943 die Volksschule in Schwerin und vom Herbst 1943 bis 1945 die Deutsche Volksschule in Brunde bei Apenrade. 1945 kam ich auf das Gymnasium in Apenrade und nach Kriegsende auf die Wissenschaftliche Oberschule für Jungen St. Georg in Hamburg, dann 1951 auf die Schleeschule in Hamburg-Othmarschen und legte dort 1953 die Reifeprüfung ab.

Im Sommersemester 1953 begann ich mein Studium an der Universität Hamburg in den Fächern Mathematik, Physik mit dem Studienziel 'Staatsexamen für das Höhere Lehramt'. Nach dem Hören der Vorlesungen von Professor Collatz wuchs das Interesse für Angewandte Mathematik. Ich studierte im Sommer 1955 ein Semester in Wien bei Professor Hofreiter, setzte anschließend mein Studium in Hamburg fort und wurde 1957 am Institut für Angewandte Mathematik als Übungsleiter und später als Hilfsassistent beschäftigt. Das Philosophikum zum Staatsexamen bestand ich im Sommersemester 1957 bei den Herren Professoren Leese und Geisler und das Staatsexamen im Sommersemester 1959 bei den Herren Professoren Collatz, Raether und Lehmann mit einer schriftlichen Arbeit über Lösungsschranken bei Randwertaufgaben elliptischer Differentialgleichungen.

Darauf wurde ich Wissenschaftlicher Mitarbeiter und zeitweilig Verwalter einer Assistentenstelle bei Herrn Professor Collatz, dem ich auch das Thema der vorliegenden Arbeit verdanke. Seit dem Frühjahr 1961 arbeite ich am Institut für Angewandte Mathematik an einem Forschungsauftrag der Deutschen Forschungsgemeinschaft über Iterative Methoden zur Lösung linearer Gleichungssysteme. Die Ergebnisse dieses Auftrages wurden in meiner Dissertation niedergelegt.

Herstellung:
XEROGRAPHIE und ROTAPRINT- Offset
G M L Wittenborn Söhne Hamburg 13